



Three Essays on Macroeconomics

Kirill Shakhnov

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence, 29 April 2015

European University Institute
Department of Economics

Three Essays on Macroeconomics

Kirill Shakhnov

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Árpád Ábrahám, EUI, Supervisor

Prof. Mark A. Aguiar, Princeton University

Prof. Rodolfo Manuelli, Washington University in St. Louis and Fed St. Louis

Prof. Ramon Marimon, EUI

© Kirill Shakhnov, 2015

No part of this thesis may be copied, reproduced or transmitted without prior
permission of the author

Abstract

This dissertation contains two lines of research: the allocation of talent and development; and sovereign default.

The first chapter contributes to the policy debate on whether the rapid growth of the US financial sector is socially desirable. I propose a heterogeneous agent model with asymmetric information and matching frictions that produces a trade-off between finance and entrepreneurship. By becoming bankers, talented individuals efficiently match investors with entrepreneurs, but do not internalize the negative effect on the pool of talented entrepreneurs. Thus, the financial sector is inefficiently large in equilibrium, and this inefficiency increases with wealth inequality. The model explains the simultaneous growth of wealth inequality and finance in the US, and why more unequal countries have larger financial sectors.

The second chapter explains the simultaneous growth of the services sector and income inequality by studying an endogenous educational choice of heterogeneous agents in the form of talent. There are two mechanisms of financing higher education: bequests and loans. The model with bequests predicts an endogenous and permanent separation of the population between the rich and the poor. The model with loans allows for social mobility, but still generates a persistent level of inequality. On the transition from the traditional economy with bequests to the economy with loans, the model qualitatively reproduces the dynamics of skill supply, the college wage premium, tuition fees and the labor allocation between sectors in the last century in the US.

The third chapter provides a novel theory to explain why sovereigns borrow on both domestic and international markets and why defaults are mostly selective (on either domestic or foreign investors). Domestic debt issuance can only smooth tax distortion shocks, whereas foreign debt can also smooth productivity shocks. If the correlation of these shocks is sufficiently low, the sovereign borrows on both markets to avoid excess consumption volatility. Defaults on both types of investors arise in equilibrium due to market incompleteness and the government's limited commitment. The model matches business cycle moments and frequencies of different types of defaults in emerging economies. We also find, contrary to existing

contributions, that secondary markets are likely to increase the risk of sovereign defaults. The outcome of the trade in bonds on secondary markets depends on how well each group of investors can coordinate their actions.

Acknowledgments

This section is dedicated to the people who helped me on the way to my Ph.D. I would like to express my special appreciation and thanks to my supervisors Árpád Ábrahám and Ramon Marimon. Thanks to Árpád, for all your efforts and guidance in my research and beyond. I am extremely grateful for your limitless support, especially during the job market – you are a terrific mentor. Thank to Ramon, for your dedication and faith in me. This thesis would never have been completed without your efforts. I am also extremely grateful to all members of the economics department for providing exceptionally well training. I thank Evi Pappa and Juan Dolado for countless hours of discussion.

I would also like to thank my coauthor of the third chapter and friend, Wojtek Paczos, for his tremendous contribution to the chapter and my social life. Many people and friends have helped, supported and shared the experience with me during my Ph.D. I am indebted to Srećko Zimic, Romanos Priftis, Vasja Sivec, Andrew Gimber, Marek Raczko, Alastair Ball, Jeanine Baumert, Antoine Camous, Matteo Fiorini, Abián García Rodríguez, Anna Gibert Rivas, Adam Jakubik, Mathilde Lebrand, Carlos Nogueira (Gama), Reinhard Ellwanger, Øystein Hernæs, Marek Raczko, Kyriakos Sotopoulos and Mehmet Emre Tokgoz.

I would also like to thank the administrative staff at the Economics department, in particular, Thomas Bourke, Riccardo Mongiu, Jessica Spataro, Julia Valerio and Lucia Vigna for their always prompt help with many issues.

I am particularly thankful to my parents who constantly supported and inspired me. Finally, I am dedicating this Ph.D. to my wife, who made all this possible, and my daughter, who motivated me.

Introduction

This dissertation contains two lines of research that I have been pursuing during the course of my Ph.D. the allocation of talent and development; and sovereign default. A common theme among them is to develop a macroeconomic theory of economies with frictions, while addressing policy issues I consider relevant. In what follows I provide a brief summary of these lines of research.

The Allocation of Talent and Development

The first theme of my thesis is to analyze the effect of *ex-ante heterogeneity* and *frictions* on economic outcomes: growth, inequality, etc. The main challenge is to correctly identify the “exact” source of inefficiency. Once this has been done, we can address the normative question: what policies can correct the inefficiency? I have focused on wealth and talent as important sources of heterogeneity.

The first chapter, “*The allocation of talent: finance versus entrepreneurship*”, studies the effect of private information and matching on *the allocation of talent and wealth* in a framework with ex-ante heterogeneous agents who face an occupational choice. The paper analyzes a meaningful tradeoff between financial intermediation and production. Financial intermediation is socially desirable, because it increases the efficiency of production, but does not directly contribute to production. The allocation of talent is inefficient due to externalities: bankers do not internalize the fact that more bankers means fewer entrepreneurs. The model provides an explanation for the expansion of the finance sector and assesses the efficiency of the expansion. I argue that the unequal accumulation of wealth leads to the expansion of the financial sector. Small initial differences in wealth among investors cause substantial income inequality among entrepreneurs, which is translated into greater wealth inequality next period. Wealthy investors are willing to pay a higher premium for financial services that increase the return on their savings, and so the greater is the dispersion of wealth, the higher is the price of financial services. This higher price induces a larger fraction of talented agents to pursue careers in finance. Hence, the growth of finance and the increase in wealth inequality go hand in hand. The paper provides an novel explanation for why the financial sector might grow too large.

The second chapter, “*Structural changes and labor income distribution: the importance of educational policies*”, looks at the allocation of talent between services and manufacturing. This paper explains the simultaneous growth of the services sector and income inequality by studying the impact of an endogenous educational choice on both sector composition and labor income distribution. In my analysis, there are two mechanisms of financing higher education: bequests and loans. The model with inheritance predicts an endogenous and permanent separation of the population between the rich (skilled) and the poor (unskilled), and generates a misallocation of talent. The model with loans leads to an efficient outcome, even though it still may generate a persistent level of inequality. Putting both models into a historical perspective, considering bequests as the traditional way to finance higher education, and loans as the modern one, the economy, switching from bequests to loans, qualitatively reproduces the dynamics of the supply of skills, the college wage premium, tuition fees and the labor allocation between sectors in the last century in the US. This novel explanation has not previously been studied in the literature.

Sovereign Default

The second theme, strategic sovereign default, has been studied extensively in the literature. The literature seeks to explain the incentives for sovereign borrowers to *repay their debts*, and hence also the incentives for creditors to lend to sovereigns in the first place. In addition, sovereign default models have proven to be a useful tool to understand and predict the dynamics of macroeconomic variables.

The third chapter, joint work with Wojtek Paczos (EUI), “*Sovereign debt issuance and selective default*”, looks at *two types of debts*: external and domestic, and how debt composition changes government incentives to borrow and default selectively. Domestic debts and, consequently, domestic defaults have been neglected in the literature despite the fact that empirical evidence shows that in the vast majority of cases, governments default selectively either on domestic or external public debt holdings, and at least 58 de jure sovereign defaults on domestic public debt have happened over the last century. As the question why governments usually default selectively on either foreign or domestic debt remains

open, this paper is an attempt to fill this gap.

We consider standard frictions: incomplete markets and limited commitment of the government. The government has to cover its expenditures by using three sources of funding: distortionary taxes, domestic debt and foreign debt. The government uses external borrowing to smooth output fluctuations. Consequently, consistent with the strategic sovereign default literature, external default is more likely in recessions, when a risk-averse borrower finds it more costly to repay non-contingent debt. The government finds it optimal to use mostly domestic debt to smooth distortions. Domestic default is thus more likely when tax distortions are high. Total default happens when high distortions coincide with a recession. Second, the model matches important data moments: a reasonably high debt-to-GDP ratio with a reasonably low default probability. Third, contrary to recent theoretical findings we show that trade in secondary markets might increase the risk of sovereign default.

Contents

Chapter 1	The Allocation of Talent: Finance versus Entrepreneurship	1
1.1	Introduction	1
1.2	Static model	6
1.2.1	Environment	7
1.2.2	Simple model without finance	9
1.2.3	First best vs. random matching	10
1.2.4	The role of finance in the model	12
1.2.5	Constrained efficiency	15
1.2.6	Decentralized equilibrium	17
1.2.7	Taxation	23
1.2.8	Comparative statics	24
1.3	Dynamic model and quantitative analysis	25
1.3.1	Dynamic model	25
1.3.2	The US experience	29
1.4	Cross-country data	33
1.5	Discussion and conclusion	36
1.5.1	Conclusion	39
1.6	Appendix A	47
1.6.1	Proof of Proposition 1	47
1.6.2	Proof of Proposition 2	49
1.6.3	The solution of the decentralized equilibrium if $\gamma^* = 0$	53
1.7	Appendix B	54

1.7.1	Private banking and private equity finance	54
Chapter 2 Structural Changes and Labor Income Distribution: The Importance of Educational Policies		57
2.1	Introduction	57
2.2	Related Literature	60
2.3	Data	63
2.4	Model	68
2.4.1	Production	69
2.4.2	Model with Bequests	71
2.4.3	Model with Loans	74
2.4.4	Income Inequality	76
2.5	Simulations and Results	76
2.6	Conclusion	78
2.7	Appendix	85
Chapter 3 Sovereign Debt Issuance and Selective Default		103
3.1	Introduction	103
3.2	Facts	107
3.3	Two-period model	111
3.3.1	Default schedule	113
3.3.2	Debt policies in the first period	118
3.4	Quantitative analysis	120
3.4.1	Households	120
3.4.2	Foreign investors	121
3.4.3	Recursive equilibrium	121
3.4.4	Calibration	125
3.4.5	Simulation results	127
3.5	Secondary markets and haircuts	132
3.5.1	Setup	132
3.5.2	Equilibria when foreign debt overhang dominates	138
3.5.3	Equilibria when domestic accommodation space dominates	142
3.6	Conclusions	143

3.7	Appendix	147
3.7.1	Two-period model	147
3.7.2	Solution algorithm	148
3.7.3	Two-period model (graphical solutions)	150
3.7.4	Secondary market: proposition proofs	151

Chapter 1

The Allocation of Talent: Finance versus Entrepreneurship

1.1 Introduction

“We are throwing more and more of our resources, including the cream of our youth, into financial activities remote from the production of goods and services, into activities that generate high private rewards disproportionate to their social productivity.”

— James Tobin (1984)

The growth of the financial sector is well known and well documented. Figure 1.1 shows that the share of finance in GDP as well as employment has increased substantially since the Second World War. The figure shows that finance accounts for a higher share of GDP than of employment before the Second World War and after the 1980s (Philippon and Reshef, 2012). More importantly, while the share of finance in employment has stabilized after the 1980s, the share of finance in GDP has continued to rise.

The substantial expansion of the financial sector has driven a debate on whether this expansion is socially desirable. On the one hand, the former chairman of the Federal Reserve, Alan Greenspan (2002) stated: “[M]any forms and layers of financial intermediation will be required if we are to capture the full benefit of our

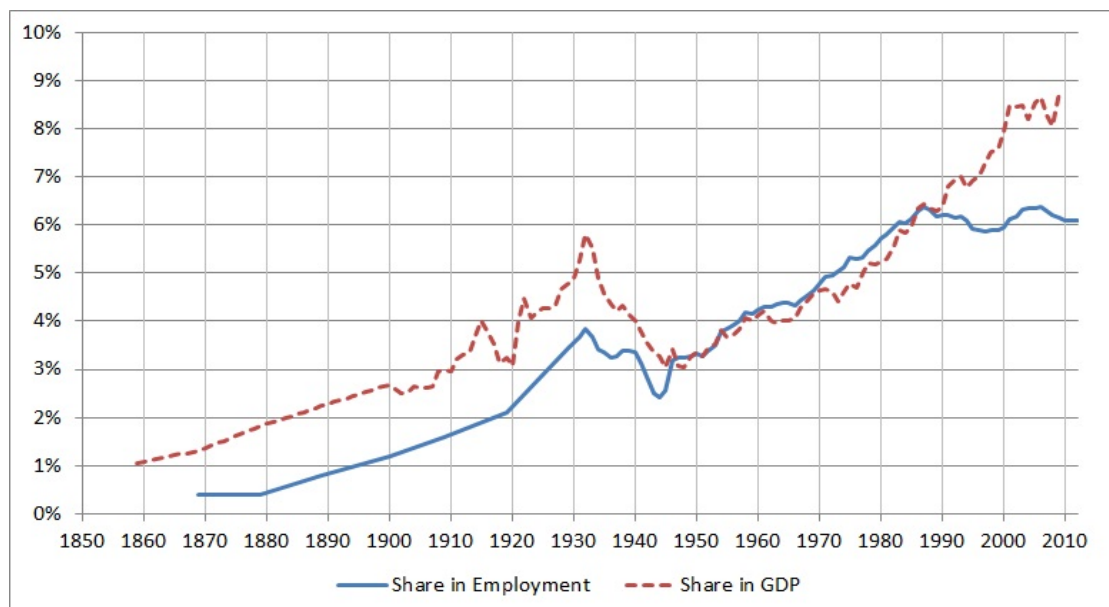


Figure 1.1: The growth of the financial sector in the US

advances in technology and trade.” This idea is related to a vast literature arguing that financial development causes economic growth, because by relaxing financial constraints the financial sector corrects capital misallocation and consequently mitigates productivity losses from financial frictions. (See Schumpeter (1934) for an early contribution and also Merton (1995). Brunnermeier et al. (2012) review the macroeconomic implications of financial frictions, while Levine (2005) provides a survey of an even larger empirical literature.)

On the other hand, critics of the financial sector suggest that it might have negative implications for the allocation of talent. Another former chairman of the Federal Reserve, Paul Volcker (2010) clearly stated the issue: “[I]f the financial sector in the United States is so important that it generates 40% of all the profits in the country... What about the effect of incentives on all our best young talent, particularly of a numerical kind, in the United States?” Furthermore, this concern has been vividly expressed on both sides of the Atlantic, in particular by Lord Turner, the former chairman of the UK’s Financial Services Authority, who stated in 2009 that the financial sector had increased “beyond a socially reasonable size.” Barack Obama (2012) and James Tobin (1984) tend to agree. This concern has

been supported by empirical findings. For example, Berkes et al. (2012) suggest that finance starts having a negative effect on output growth when credit to the private sector reaches 100% of GDP. Other authors, such as Goldsmith (1995) and Lucas (1988), claim that the role of finance has been overstated, and argue that it responds passively to economic growth.

In order to evaluate these claims in a structured way, I build a model in which financial intermediation potentially enhances welfare but draws some talented individuals away from production. The model includes three key elements: (a) heterogeneous agents who differ in terms of capital and talent; (b) an occupational choice between being a banker or an entrepreneur; (c) financial frictions. Heterogeneity and an occupational choice provide a framework to study the allocation of capital (wealth) and talent. Talent is important for both industry and the financial sector: more talent in industry means more output is produced per unit of capital, while more talent in finance means capital is potentially allocated more efficiently. Financial frictions in the form of private information cause the misallocation of capital, because investors cannot distinguish between talented and ordinary entrepreneurs. Since talented bankers can make this distinction, the financial sector can potentially correct this misallocation.

The model generates four important insights about the financial sector. First, it implies that the optimal (constrained efficient) size of the financial sector is larger for countries or periods with higher wealth and talent inequality, because in these cases the potential productivity losses from capital misallocation are particularly severe. The planner faces a tradeoff between the misallocation of capital and the misallocation of talent. Second, the decentralized equilibrium exhibits a misallocation of talent: the financial sector absorbs talent beyond the socially desirable level, because it provides talented agents with an opportunity to extract an excessive informational rent due to the presence of externalities. When agents make their occupational choice between finance and entrepreneurship, they do not internalize the negative externality that they impose on investors: the more bankers there are, the fewer talented entrepreneurs and good investment opportunities there are. Third, even though the equilibrium is generically *inefficient*,

efficiency can be restored by taxing the financial sector. Fourth, the model provides a novel explanation for the growth of finance by linking it to an increase in *wealth inequality*. In the dynamic framework, this effect is self-reinforcing: small initial differences in wealth among investors cause substantial income inequality among entrepreneurs, which is translated into greater wealth inequality next period. Wealthy investors are willing to pay a higher premium for financial services that increase the return on their savings, and so the greater is the dispersion of wealth, the higher is the price of financial services. This higher price induces a larger fraction of talented agents to pursue careers in finance. Hence, the growth of finance and the increase in wealth inequality go hand in hand.

Some papers provide indirect empirical evidence on the misallocation of talent. Data from college graduates in the US suggests that the financial sector has become one of the most popular destinations for graduates of elite universities, regardless of their major. For example, Shu (2012), studying the career choices of MIT graduates, concludes that careers in finance attract students with high levels of raw academic talent. She concludes that the overall allocation of talent is inefficient. (See also Goldin and Katz (2008) for Harvard graduates, and Wadhwa et al. (2006) for Engineering Management graduates at Duke University.) In addition, Kneer (2012) finds that US banking deregulation reduces labor productivity disproportionately in industries that are relatively skill-intensive. Finally, MGI (2011) estimates that the United States may face a shortfall of almost two million technical and analytic workers over the next ten years.

This paper is related to a vast literature on misallocation, particularly to papers attributing the misallocation of capital to financial frictions Buera and Shin (2013); Midrigan and Xu (2014). Whereas most papers focus on the impact of frictions on output and the allocation of capital, and abstract away from its impact on the labor market and the allocation of human capital (Jovanovic (2014) is one of the exceptions), this paper argues that financial development has an important impact on the allocation of both capital *and* talent, which cannot be neglected. The issue of allocative efficiency has also been studied theoretically in relation to venture capital. For example, Jovanovic and Szentes (2013) show that the competitive

equilibrium is always socially optimal, while in search-matching models such as Michelacci and Suarez (2004), the Hosios condition must hold for the equilibrium to be efficient.

Apart from the current paper, three recent papers have analyzed whether the expansion of the financial sector is efficient. The financial sector is inefficient in all three papers, but the source of the inefficiency is different. Murphy et al. (1991) argue that the flow of talented individuals into law and finance might not be entirely desirable, because even though private returns in these occupations are high, social returns might be higher in other occupations. However, they provide no reason for the disparities between social and private returns. The study of Philippon (2010) is the first that acknowledges the meaningful role of the financial sector, a monitoring device that helps to overcome the opportunistic behavior of entrepreneurs. The allocation is not optimal in his model, because the projects developed by entrepreneurs have higher social benefits than private ones; therefore, they need to be subsidized with respect to workers and bankers. Bolton et al. (2011) focus on financial innovations, in the sense that the financial sector creates a new over-the-counter (OTC) market. Informed dealers in the OTC market extract excessive rents, and consequently the financial sector attracts too many individuals. However, none of these papers seek to explain the growth of the financial sector; none of them consider the financial sector as financial intermediaries connecting investors and entrepreneurs; neither Murphy et al. (1991) nor Philippon (2010) allow for excessive informational rent extraction; and finally, neither Philippon (2010) nor Bolton et al. (2011) have a role for talent in either finance or industry.

Many studies analyze the causes of the expansion of the financial sector. Several explanations have been suggested: the fluctuation of trust in financial intermediaries Gennaioli et al. (2013); the increasing efficiency of the production sector Bauer and Mora (2014); structural change in finance Cooley et al. (2013); and asset bubbles Cahuc and Challe (2012). None of them connect the expansion of the financial sector and the increase in wealth inequality. The only paper that partially attributes the growth of finance to capital accumulation is Gennaioli et al. (2013). I focus not on aggregate capital accumulation, but rather on increasing

wealth inequality. I show that the growth of wealth inequality alone is enough to fully explain the growth of finance. This is in line with Piketty and Zucman (2014)'s argument that the primary reason for increased inequality is the fact that financial services associated with asset management generate superior returns and disproportionately affect the wealthy. According to Greenwood and Scharfstein (2013), much of the growth of the financial sector comes from asset management, which is mostly a service for wealthy individuals.

The calibrated model qualitatively replicates well other features of the US data: the increase in wealth inequality, the productivity slowdown, and the growth of the financial sector as a share of both employment and GDP. The model predicts that the financial sector would continue to grow as a share of GDP, but not of employment. It also provides an additional explanation for the US productivity slowdown. Furthermore, cross-country regressions show that, in line with the predictions of the model, inequalities of wealth and talent are positively associated with the size of the financial sector.

The paper is structured as follows. Section 1.2 describes the static version of the model and policy results. Section 1.3 provides the dynamic version of the model and quantitative analysis. Section 1.4 performs a cross-country analysis to confirm the findings. The last section discusses the paper, concludes, and motivates further research.

1.2 Static model

There are two opposing views on finance. On the one hand, a large literature on finance and development establishes a positive link between finance and aggregate output. From the theory side, the standard way to think about the issue is that, due to financial frictions, there is misallocation of capital and consequently output losses, which can be severe. The financial sector plays an important role in overcoming or at least mitigating the effect of these frictions. Based on this view, the main policy prescription is to promote the development of the financial sector. On the other hand, the Great Recession has cast doubt on the efficiency of

the rapid growth of finance, suggesting that possible rent seeking behavior might be involved. The model presented below features financial frictions that generate capital misallocation. The financial sector can correct this misallocation at the cost of talent misallocation.

I adopt the “classical” view of financial intermediaries as institutions that connect surplus agents (investors) and deficit agents (entrepreneurs). Financial intermediaries are efficient at obtaining information, but they require talent to acquire this information. A talented banker can screen entrepreneurs to discover the best investment opportunities, and sells this information to an investor. The financial sector in the model is clearly a productive sector, because it mitigates informational frictions.

1.2.1 Environment

The economy consists of two types of agents: *investors and entrepreneurs*. To produce output, two inputs are required: capital and an idea. Investors have wealth but no investment opportunities of their own, while entrepreneurs have ideas but need external funding. The *Cobb–Douglas production function* is assumed $F(z, k) = z^{\alpha_z} k^{\alpha_k}$.

Agents are *heterogeneously endowed with talent and wealth*. (Since capital is the only asset in the economy, the terms “wealth” and “capital” are used interchangeably.) Investors can be capital-abundant or capital-scarce, while entrepreneurs can be talented or ordinary. Entrepreneurs can *choose* whether to remain entrepreneurs or to become bankers instead. In industry, talent translates into capital productivity. The more talented is the entrepreneur, the more output is produced from a unit of capital. In finance, talent affects bankers’ ability to distinguish between talented and ordinary entrepreneurs or to sort them, as we shall see below.

I consider a two-sided *one-to-one matching market*: one entrepreneur needs to be matched with one investor to produce. The economy is subject to financial frictions: *two-sided private information*, meaning that the types of entrepreneurs (investors) are not publicly observable. When investors are looking for investment opportunities, they do not know whether an entrepreneur that they meet is tal-

ented or ordinary. The same holds for entrepreneurs: entrepreneurs do not know whether an investor they are dealing with is capital-abundant or -scarce. Even though the latter assumption seems questionable at first, in the venture capital industry it is common for entrepreneurs to be imperfectly informed about the total wealth of investors.¹ Two-sided private information guarantees that the outcome is random matching in the case of continuous distribution over types, because it is impossible to write an enforceable contract based on only one observable outcome for two unobservable inputs. In the case of discrete distributions, we need to be sure that two different pairs of inputs lead to the same output ($F(z^H, k^L) = F(z^L, k^H)$). The literature on assortative matching states that as long as the private information is one-sided, there is a separating equilibrium that supports the same positive assignment as in the full-information equilibrium assignment. In the economy with private information, but without matching, the aggregate outcome is exactly as in random matching, because investors optimally allocate equal shares to every entrepreneur. Matching simply ensures that all funds are not allocated to one entrepreneur. Alternatively, we can simply assume that *without financial intermediation*, the investment technology in the economy is random matching.

All agents are assumed to be *risk-neutral and discount the future at a zero rate*, so all agents maximize their incomes and aggregate output is welfare criteria.

¹In the model, the wealth of investors is invested fully; immediately afterwards, a one-time investment output is produced. In reality, it is more complicated. Even after engaging with a venture capitalist, the entrepreneur faces a substantial degree of uncertainty about the total amount of investment, because of staging. Staging is one of the central incentive mechanisms used in the venture capital industry (Sahlman, 1990). As shown by Bienz and Hirsch (2011), staging is frequently implemented through multiple negotiated financing rounds. Furthermore, the venture capital literature often assumes that *neither* the inputs of the investor nor those of the entrepreneur are contractible. The standard feasible contract in the venture capital literature specifies only a sharing rule and an initial investment, but not the total investment, which, like entrepreneurial inputs, is assumed to be noncontractible.

1.2.2 Simple model without finance

This subsection presents a simple static general equilibrium model with unobserved heterogeneity. The model without finance and full information is a variant of the standard static model of two-sided matching in which a Becker–Brock type of assignment problem arises (Becker, 1973). I add to this framework two features: two-sided private information and intermediation. Two-sided private information ensures that the assignment should be random—*without intermediation* (the financial sector), there is no mechanism to enforce positive assignment (assortative matching). The full dynamic model presented in the next subsection will incorporate this same static model into a dynamic framework.

In this section, for the sake of simplicity, I consider a *very particular distribution of wealth and talent*: there is a unit mass of agents with talent and no capital, who can be talented z^H or ordinary z^L ; there is a unit mass of agents with capital and no talent, who can be capital-abundant k^H or capital-scarce k^L . The share of capital-abundant investors (talented entrepreneurs) is denoted as β^i (β^e). Hence, the mass of agents with capital is equal to the mass of agents with talent. Agents with capital and no talent are potential investors, while agents with no capital and talent can be either entrepreneurs or bankers. Every investor can be matched with at most one entrepreneur. Hence, I consider the simplest case of matching, which is *one-to-one matching*. Furthermore, I assume that all short-sided agents are matched with certainty.² The outcome of the match is given by a *strictly supermodular function* $F(z, k)$ depending on both capital and talent. The strict supermodularity in the discrete case is given by:

$$F(z^H, k^H) + F(z^L, k^L) > F(z^H, k^L) + F(z^L, k^H) \quad (1.1)$$

Condition (1.1) suggests that positive assortative matching maximizes the sum of match outputs when the entrepreneur's type and the investor's type are complements in the match output function.

²One-to-one matching can be viewed as a technological constraint. Many-to-one matching, different specifications of the matching function and a continuum of types over talent and wealth are discussed in section 1.5 below.

Figure 1.2 shows the outcome of matches in this economy. Since investors and entrepreneurs can be of only two types, we have four possible outcomes (sky blue, yellow, pink and orange). I introduce an additional notation $F_{IJ} = F(z^I, k^J)$, where $I, J = \{H, L\}$, I stands for the entrepreneur's type and J stands for the investor's type. For example, F_{HL} is the outcome of a match between a high-type entrepreneur and a low-type investor; the yellow area is the combination of two colors: green z^H and brown k^L .

		Investors	
		Abundant k^H	Scarce k^L
Entrepreneurs	Talented z^H	β_i $F(z^H, k^H)$	$1-\beta_i$ $F(z^H, k^L)$
	Ordinary z^L	$1-\beta_e$ $F(z^L, k^H)$	$F(z^L, k^L)$

Figure 1.2: Model without bankers

1.2.3 First best vs. random matching

In this section, I define the *first best* as an optimal allocation under the constraint of the matching technology. Since the financial sector mitigates information frictions but does not directly contribute to production, the first best in this economy is the allocation in which nobody is a banker and all talented agents are matched with investors. Under the supermodularity assumption on the production function (outcome of the match) and observability of types, the most efficient

outcome in this economy is positive assortative matching—when high-type entrepreneurs are matched with high-type investors, and low-type entrepreneurs are matched with low-type investors (see the Becker–Brock efficient matching theorem). However, in the case when $F_{HL} = F_{LH}$ assortative matching cannot be achieved due to two-sided private information about types, so I consider the assortative matching outcome as the first-best allocation. The only possible outcome in the economy with private information and without a financial sector is random matching.

The simple example below shows the disparity between the first best and random matching: the loss of aggregate output due to the misallocation of capital caused by private information in this economy can be severe. I consider the case in which the production function is simply the product of two inputs $F = zk$. I assume that the value of the high type is one with probability one-quarter, while the value of the low type is zero with the complementary probability for both the distribution of talent and the distribution of wealth. Hence, only if two high types are matched is any output (one unit) produced. It happens with probability $1/16$ in the case of random matching and with probability $1/4$ in the case of assortative matching (the first best). Table 1.1 summarizes the information described above. As we can see, output is four times lower in the case of random matching com-

Table 1.1: The simple example

	value	probability
z^H	1	1/4
z^L	0	3/4
k^H	1	1/4
k^L	0	3/4
Random matching	1/16	
Assortative matching	1/4	

pared to the first best due to capital misallocation. This brings us to the first main

question of whether the financial sector can mitigate this capital misallocation.

1.2.4 The role of finance in the model

The financial sector clearly provides many useful functions to the economy, as discussed in section 1.5. This paper focuses on two services: intermediation and sorting between investors and entrepreneurs. It is important to remember that the most desirable outcome is assortative matching. All investment goes to industry. Bankers are good at sorting, but they do not directly produce any output. The quality of sorting depends on talent. Both finance and industry require talent. While talent in industry increases the firm's productivity, talent in finance gives an *advantage in obtaining information* and therefore increases the quality of sorting. By the latter, I mean that the financial sector brings the allocation as close as possible to the allocation under assortative matching. However, the allocation under assortative matching cannot be achieved. As a reminder, the first-best allocation is the allocation under assortative matching; the allocation without financial intermediation is the allocation under random matching. I call the allocation with financial intermediation the allocation under *intermediated matching*. It is important to distinguish between the constrained efficient allocation under intermediated matching, discussed in the next subsection, and the decentralized allocation under intermediated matching, discussed later.

For most of this paper, if not specified otherwise, I consider an extreme case in which only the high-type z^H banker can match a talented entrepreneur and a capital-abundant investor for sure, while the low-type z^L banker can only match randomly. This assumption has two possible interpretations. Under the first interpretation, the quality of sorting depends on the talent of the agent who does the sorting. A banker with ability z can distinguish between ideas with productivity z and $z' < z$. Hence, the planner would only consider allocating talented z^H agents to finance.

Under the second interpretation, there is a cost of screening $\psi(z)$ per project discovered, which depends on talent. If this cost is high enough for the low type while low enough for the high type, $\psi(z^L) \gg \psi(z^H)$, then the planner might

find it optimal to allocate to intermediation some of the talented agents, who can provide efficient matches at a small cost, while she would not allocate any of the ordinary agents to intermediation because of their higher matching costs. In other words, the financial sector provides a useful service (sorting) because it has an information advantage, but requires talent to realize this advantage. This accords with Philippon and Reshef's (2012) empirical observation that working in a world of innovative finance requires talent.³

Even though the model can be applied to the financial sector as a whole, private equity finance is a subindustry for which the assumptions of the model are particularly valid: matching and information superiority. A private equity fund precisely does matching between a few selected young firms and high-net-worth individuals. The private equity fund provides an opportunity to invest in a few companies over a long-term horizon for a small number of wealthy investors (You can find more details in Appendix 1.7.1). Information superiority of the financial sector with respect to is a fairly standard assumption in finance literature supported by empirical evidence (Durnev et al., 2004; Luo, 2005; Chen et al., 2006). Furthermore, this paper abstracts from a potentially interesting extension, the trustworthiness of bankers, because the social planner can always punish bankers for an undesirable outcome in the case of intermediated matching, and it is always possible to write a contract between a banker and an investor/entrepreneur, which insures truth-telling.

I introduce an additional technical assumption: limited capacity. A banker has no capacity advantage in comparison with ordinary investors. Each banker can only provide transaction support for one deal at a time. This assumption is to ensure that one banker cannot undo all private information frictions. This assumption is discussed in detail in section 1.5.

To sum up, the two assumptions imply that if the share γ of talented agents β^e is allocated to the financial sector, they can match at most $\gamma\beta^e$ talented en-

³In other words, talented bankers provide an investment opportunity with a superior return because of their informational advantage. We can also think of agents as having different search costs in the case of search frictions.

		Investors	
		Abundant k^H	Scarce k^L
Entrepreneurs		$\gamma\beta_e$	$\frac{(\beta_e - \gamma\beta_e)}{(1 - \gamma\beta_e)}$
Talented z^H	$\gamma\beta_e$	$F(z^H, k^H)$	
Bankers	$\gamma\beta_e$		
Talented z^H	$(1 - 2\gamma)\beta_e$	$F(z^H, k^H)$	$F(z^H, k^L)$
Ordinary z^L	$1 - \beta_e$	$F(z^L, k^H)$	$F(z^L, k^L)$

Figure 1.3: Model with bankers

trepreneurs. To be precise, $\min\{\gamma, 1 - \gamma\}\beta^e$ talented entrepreneurs are matched by bankers with capital-abundant investors and $\max\{1 - 2\gamma, 0\}\beta^e$ are left for random matching. Figure 1.3 summarizes the situation stated above and describes the outcome of matches in the case of $\gamma \leq 1/2$. It is very similar to Figure 1.2, but with the addition of the financial sector. Out of talented agents β^e , the share γ is allocated to the financial sector, while the share $1 - \gamma$, together with all ordinary agents $1 - \beta^e$, is allocated to industry. We observe losses (the white area) because some investors remain unmatched, and gains (the sky blue area) because the number of efficient matches increases.

1.2.5 Constrained efficiency

In this subsection, I introduce the notion of constrained efficiency. A social planner faces the same private information constraints as individuals do. To overcome these constraints, the planner can choose consumption of agents based on observables (the number of bankers and the outcomes of matches) to make sure that a fraction of talented agents self-select themselves into the financial sector. Since only talented agents z^H have an informational advantage and can distinguish between good and bad projects, they are the only agents that need to be considered as possible bankers, because the social planner can always punish a ordinary banker. If ordinary agents z^L select to be a banker, they have no informational advantage and, hence, have a positive probability to match an ordinary entrepreneur with a capital-abundant investor, that generates the lower outcome ($F_{LH} < F_{HH}$). Since the outcome of the match is observable, the social planner assign the consumption of an ordinary banker lower enough to make sure that none of ordinary agents select themselves into finance. For simplicity, I assume that the number of investors is always greater than the number of bankers.⁴ Hence, some investors are matched with nobody. By allocating the fraction γ of talented agents to finance, the planner gains the value of intermediated matches between talented entrepreneurs and capital-abundant investors F_{HH} and incurs two costs: the direct

⁴I prove that this is necessary for the existence of a decentralized equilibrium. See the proof of Proposition 2 in appendix 1.6.2

cost is due to the fact that $\gamma\beta^e$ investors become unmatched; the indirect one is that the probability of being randomly matched with talented entrepreneurs drops substantially. Because of risk neutrality, the constrained efficient allocation is one that maximizes aggregate output. The precise expression for aggregate output is given by

$$\mathbf{Y} = \max_{\gamma} \left\{ \frac{(\beta^i - \min\{\gamma, 1-\gamma\}\beta^e)}{(1 - \min\{\gamma, 1-\gamma\}\beta^e)} [\max\{1 - 2\gamma, 0\}\beta^e F_{HH} + (1 - \beta^e)F_{LH}] + \frac{(1 - \beta^i)}{(1 - \min\{\gamma, 1-\gamma\}\beta^e)} [\max\{1 - 2\gamma, 0\}\beta^e F_{HL} + (1 - \beta^e)F_{LL}] + \min\{\gamma, 1 - \gamma\}\beta^e F_{HH} \right\}. \quad (1.2)$$

As soon as γ exceeds $1/2$, all talented entrepreneurs are matched with capital-abundant investors. There is no gain to allocating an additional talented agent to the financial sector. Therefore, the constrained efficient allocation γ^* cannot exceed $1/2$; otherwise we would observe pure losses in the quantity of talented entrepreneurs without any additional gains from matching, which cannot be efficient.

Proposition 1 describes the solution of problem (1.2):

Proposition 1. *The constrained efficient allocation γ^* is always the corner solution of problem (1.2), i.e. γ^* can be either 0 or $1/2$.*

Proof. See Appendix 1.6.1.

I calculate $\Delta\mathbf{Y}$, the difference between the values of the planner's objective (1.2) with $\gamma = 1/2$ and $\gamma = 0$. This difference is given by

$$\Delta\mathbf{Y} = (1/2 - \beta^i)\beta^e(F_{HH} - F_{HL}) - \frac{\beta^e}{2}F_{HL} - \frac{(1 - \beta^i)(1 - \beta^e)\beta^e}{2 - \beta^e}(F_{LH} - F_{LL}). \quad (1.3)$$

After analyzing expression (1.3) above, we can conclude that if $\beta^i \geq 1/2$, $\gamma^* = 0$ is the only possible solution of the planner's problem. For $\gamma^* = 1/2$ to be the solution, two conditions must be satisfied: $\beta^i < 1/2$, and F_{HH} needs to be high enough. In other words, it is efficient to have a financial sector if two requirements are met: the probability of a random match between a talented entrepreneur and a capital-abundant investor is relatively low, but the value of this match is relatively high. I provide two potential interpretations of this result. On the one hand, one might think that the level of development affects the optimal size of the financial sector. In a developing country with weak institutions, it is difficult for an investor

to meet the “right” entrepreneur. Hence, it is essential for such countries to develop their financial sectors to mitigate the effect of underdeveloped institutions. The conclusion might be that the more developed a country is, the less likely it is to benefit from the financial sector. This conclusion seems at best to be counterfactual. However, Mayer-Haug et al. (2013) observe that entrepreneurial talent is more relevant in developing economies. Furthermore, empirical evidence suggests that the misallocation of capital is a particularly acute problem in developing countries. On the other hand, having a financial sector is efficient for countries with higher degrees of wealth or talent inequality. The more unequal a country is, the higher are the benefits from the presence of the financial sector. I provide empirical support for the latter interpretation in section 1.4. (See also Restuccia and Rogerson (2008) for the argument that resource misallocation shows up as low TFP, and Hsieh and Klenow (2009) for empirical evidence on misallocation in China and India.)

If we go back to the simple example in Table 1.1 and calculate aggregate output in the constrained efficient case, we obtain $1/2\beta^e F_{HH} = 1/8$, which is twice as large as in the case of random matching (the economy without finance), but still two times lower than in the first best. In the case of the simple example, we can say that the financial sector undoes half of the financial friction.

1.2.6 Decentralized equilibrium

In this subsection, I study the decentralized equilibrium (DE) and compare it to the constrained efficient allocation to answer the question of whether the financial sector attracts the right amount of talent. The main difference between the DE and the constrained efficient allocation is the fact that the occupational choice of agents depends on the *private* returns in the two sectors, as opposed to *social* returns in the planner’s case. The planner chooses consumption of agents based on observables (the number of bankers and the outcomes of a match) to make sure that the right number of talented agents to self-select themselves into finance and at the same time how much consumption they should get. Given the information structure, it is a complicated task for the market to solve, because the number of talented

agents in finance affects the way the surplus is shared between three parties: an investor, an entrepreneur and a banker. On the one hand, surplus is created by agents in industry (entrepreneurs). On the other hand, private information frictions create an information rent that can be captured by agents in the financial sector (bankers). In addition, due to matching it is important to understand how the outcome of the project is split between the investor and the entrepreneur. The most natural way to do this is *Nash bargaining*, where the bargaining power of the entrepreneur $\delta \in [0, 1]$ is exogenously given, and the bargaining power of the investor is the complement $1 - \delta$.

The timing of the problem is as following. The problem is one shoot game. First, anticipating equilibrium outcomes agents choose occupations and cannot reoptimize. The talented banker screens entrepreneurs until she finds a talented one. If the banker succeeds, she signs a contract to seek exclusive representation promising to deliver a investor with a capital k^H in the exchange for fees p^e . The banker posts a contract for in exchanges for fee p^i promising to match with a talented entrepreneur z^H . If an investor and an entrepreneur agree to sign a contract with a banker, they meet and split the outcome of the match according to the entrepreneurial bargaining power δ . Then, the banker collects fees potentially from both parties p^i and p^e . Investors and entrepreneurs can always prefer to be matched randomly for free. Random matching is the outside option for investors and entrepreneurs. In addition, I study the equilibrium of occupational choice in pure strategy. Agents cannot mix to be a banker and an entrepreneur with a positive probability.

The rest remains as outlined in subsection 1.2.1. The banker with talent z^H can distinguish between entrepreneurs with productivity z^H and $z^L < z^H$. She can sell this information to an investor for price p_i and an entrepreneur for price p_e . Each talented banker can discover at most one talented entrepreneur z^H and consequently makes at most one match between a capital-abundant investor and a talented entrepreneur. If an investor (entrepreneur) pays p_i (p_e), she knows that she will be matched with a high-type counterpart with certainty; otherwise, she can always choose to be randomly matched for free. I assume that if there are

more bankers than talented entrepreneurs, $\gamma > 1/2$, some of the bankers discover nothing and therefore receive zero income. In this case, bankers bear all the risk and need to be compensated for this. Equilibrium prices are set competitively.

Returning to the Nash bargaining problem, to solve the problem, I need to define the bargaining power, the surplus of the match and the outside options of the two counterparties. The outside option to intermediated matching is random matching. Hence, the problem must be solved *backwards*. First, I provide the solution for random matching with a given size of the financial sector γ . Then, I use the solution for random matching as the outside options for the intermediated matching problem.

To solve a Nash bargaining problem following Nash (1950, 1953), I need to define the set of feasible utility payoffs from an agreement U and the utility payoffs to the players from a disagreement D . Since preferences are linear, the sets U and D are given by

$$U = \{(x^e, x^i) | x^e + x^i = F(z, k), x^j \geq 0\}, \quad (1.4)$$

$$D = \{(d^e, d^i) | \}, \quad (1.5)$$

where x^e and x^i are the payoffs to the entrepreneur and to the investor. The entrepreneur's payoff is

$$x^e = \arg \max \left[(x - d^e)^\delta (F(z, k) - x - d^i)^{1-\delta} \right]. \quad (1.6)$$

The solutions are:

$$x^e = \delta (F(z, k) - d^i) + (1 - \delta)d^e, \quad (1.7)$$

$$x^i = (1 - \delta) (F(z, k) - d^e) + \delta d^i. \quad (1.8)$$

As every banker can discover at most one good project, the total number of discovered good projects that are different from each other is $\min\{\gamma, 1 - \gamma\}$. It is worth mentioning that, contrary to the planner's solution to problem (1.2), $\gamma^* \leq 1/2$, the market outcome can be any number in the interval $[0, 1]$.

I assume that investors have no access to a storage technology, while entrepreneurs have no opportunity for outside borrowing. Thus, the outside options

for a random match—the set D in (1.5)—are $(0, 0)$. The solution of the Nash bargaining problem gives the value of random matching for a capital-abundant investor. Note that not all investors are matched. The value of random matching is equal to the probability of matching with somebody Pr^m multiplied by the sum of products of the probability of being matched with a talented (ordinary) entrepreneur Pr^H (Pr^L) and the value of the match for a capital-abundant investor $(1 - \delta)F(z^I, k^H)$. It turns out that:

$$\begin{aligned} Pr^m &= \frac{1 - \gamma\beta^e - \min\{\gamma, 1 - \gamma\}\beta^e}{1 - \min\{\gamma, 1 - \gamma\}\beta^e}, \\ Pr^H &= \frac{(1 - \gamma - \min\{\gamma, 1 - \gamma\})\beta^e}{1 - \gamma\beta^e - \min\{\gamma, 1 - \gamma\}\beta^e}, \\ Pr^L &= \frac{1 - \beta^e}{1 - \gamma\beta^e - \min\{\gamma, 1 - \gamma\}\beta^e}. \end{aligned}$$

Hence the outside option for intermediated matching is

$$d^i = \frac{1 - \delta}{1 - \min\{\gamma, 1 - \gamma\}\beta^e} [(1 - \gamma - \min\{\gamma, 1 - \gamma\})\beta^e F_{HH} + (1 - \beta^e)F_{LH}] + \frac{\gamma\beta^e}{1 - \gamma\beta^e} 0. \quad (1.9)$$

Equation (1.9) defines the value of random matching for a capital-abundant investor, which is the outside option of a capital-abundant investor when negotiating a deal with a talented entrepreneur after intermediated matching. It is important to note that an increase in the size of the financial sector γ worsens the outside option of the capital-abundant investor, because it affects the relative proportions of agents. I return to this point later on.

Similar to (1.9), the value of random matching for a talented entrepreneur, which is the outside option for bargaining in the case of intermediated matching, is

$$d^e = \frac{\delta}{1 - \min\{\gamma, 1 - \gamma\}\beta^e} [(\beta^i - \min\{\gamma, 1 - \gamma\}\beta^e)F_{HH} + (1 - \beta^i)F_{HL}]. \quad (1.10)$$

Applying once again the solution of Nash bargaining (1.7) to the intermediated matching case, I obtain the restriction on the prices that can be extracted from

investors (1.11) and entrepreneurs (1.12):

$$p_i \leq (1 - \delta)(F_{HH} - d^i - d^e), \quad (1.11)$$

$$p_e \leq \delta(F_{HH} - d^i - d^e). \quad (1.12)$$

Conditions (1.11) and (1.12) are the participation constraints of a capital-abundant investor and a talented entrepreneur. They state that both an investor and an entrepreneur being matched by a banker cannot be worse off in comparison to the random matching scenario. However, these inequalities are not necessarily binding. It depends on which agents are on the *short side* of the market. In addition, the prices obviously should be non-negative.

To complete the description of equilibria, I need an additional condition (1.13). For the solution to be interior, $\gamma \in (0, 1)$, the talented agent ($z^H > 0$) should be indifferent between being an entrepreneur or a banker. The income of a talented banker is the probability of finding a talented entrepreneur multiplied by the sum of the two prices that are charged to the investor and the entrepreneur. As long as there are more talented entrepreneurs in the market than bankers, the probability of finding a talented entrepreneur is equal to one. The income of a talented entrepreneur is the share of the surplus received from the match with a capital-abundant investor. The indifference condition is therefore

$$\frac{\min\{\gamma, 1 - \gamma\}}{\gamma}(p_i + p_e) = \delta(F_{HH} - d^i) + (1 - \delta)d^e. \quad (1.13)$$

Three conditions characterize all decentralized equilibria: the occupational choice condition (1.13) and two participation constraints in financial services, one for capital-abundant investors (1.11) and one for talented entrepreneurs (1.12). For the sake of space, I restrict my attention to the case in which the exogenous parameters are such that the constrained efficient size of the financial sector is *strictly positive* ($\gamma^* = 1/2$). I take the view that the financial sector is essential for the economy. Furthermore, this is the interesting case in which to study policy, because for regions of the parameter space in which the financial sector plays no useful role, policy analysis is trivial. Proposition 2 characterizes the decentralized equilibrium in the $\gamma^* = 1/2$ case in terms of efficiency. A detailed analysis of all possible cases can be found in appendix 1.6.2

Proposition 2. *If it is socially efficient to have a financial sector ($\gamma^* = 1/2$) and a decentralized equilibrium exists,*

- i. It is unique, $\hat{\gamma}$;*
- ii. This equilibrium is generically inefficient, $\hat{\gamma} \geq \gamma^*$; and*
- iii. There exists a restriction on the set of exogenous parameters that restores the constrained efficient allocation.*

Proof. See Appendix 1.6.2.

This restriction can be expressed as $\hat{\delta} = f(\underset{+}{\beta^e}, \underset{+}{z^H/z^L}, \underset{-}{\beta^i}, \underset{+}{k^H/k^L})$. The signs beneath the expression stand for the sign of the derivative of $\hat{\delta}$ with respect to the corresponding variable.

Part (iii) of Proposition 2 might look similar to the Hosios condition in the sense that the condition ensures the externalities cancel out (Hosios, 1990). In the original case of Hosios, efficiency is achieved when the surplus share (bargaining power) between workers and a firm is equal to the matching share (the elasticity of the matching function). In a frictionless environment, there is a particular mechanism, directed search, that restores efficiency. However, in a frictional environment with heterogeneous agents even directed search might not be sufficient. The latter might be the case of my model.

The result stated in Proposition 2 has a very intuitive explanation. When talented agents make their occupational choice between finance and entrepreneurship, they do not internalize the externalities that they impose on investors. The more talented agents become bankers, the smaller is the pool of good projects. The bargaining process, matching friction and timing are important for this result. First, a different bargaining process might incorporate more information and take into account the externality. Second, in the perfectly competitive market prices would adjust to eliminate the externality imposed by occupational choice. Third, an infinitely repeated game, when agents can constantly switch from random to intermediated matching and constantly change occupation, should converge to an efficient solution.

The opposite case, in which the set of parameters is such that the constrained efficient size of the financial sector is *zero* ($\gamma^* = 0$) is discussed in appendix 1.6.2. The model of Murphy et al. (1991) can be viewed as a special case of my model under parameter restrictions such that $\gamma^* = 0$.

Proposition 2 states that the decentralized equilibrium is generically inefficient. To put it differently, for a given set of parameters, the solution of the decentralized equilibrium is highly unlikely to be efficient. The question is whether it is possible to restore efficiency. The answer is yes. As discussed, there is a restriction on parameters that restores efficiency. If there is a policy instrument that directly affects one of the exogenous parameters, it is easy to ensure efficiency in the model. For example, if the planner could set the bargaining power of entrepreneurs to the particular value $\hat{\delta}$, it would make the decentralized equilibrium efficient. However, it is not very intuitive to think that such policies exist.

1.2.7 Taxation

The more interesting question is whether it is possible to restore efficiency using only one tax instrument. Fixing the set of parameters to values such that the decentralized equilibrium exists and is inefficient, I take the tax on the financial sector to be the available tax instrument.

The issue in this economy is that the return to finance is too high in comparison with entrepreneurship. Hence an efficient policy should decrease the return to finance and/or increase the return to entrepreneurship. The former can be done through taxation of the financial sector. The latter can be done through subsidizing entrepreneurship. Taxation of the financial sector has been a hot topic since the Great Recession, especially in the European Union.⁵ Subsidies for entrepreneurship are quite common: governments and donors spend billions of dollars subsidizing entrepreneurship training programs around the world (see, for example, Santarelli et al. (2006)).

I show how a tax τ on bankers' incomes can work. The revenue from this tax

⁵See the discussion of taxation proposals at the European Commission web page: http://ec.europa.eu/taxation_customs/taxation/other_taxes/financial_sector/index_en.htm.

is distributed by lump-sum transfers T to balance the government's budget. The last equation of system (1.14) represents the government's budget constraint. The system below characterizes the equilibrium with taxation:

$$\begin{aligned} x^e &= (1 - \delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)) + T, \\ c &= (1 - \delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)) - 2(1 - \delta)T - \tau, \\ x^e &= \frac{1-y}{y}c, \\ T &= \gamma\beta^e\tau. \end{aligned} \tag{1.14}$$

Given the constrained efficient level $\gamma^* = 1/2$, I impose that $\gamma = \gamma^*$ and calculate the corresponding tax rate. The solution of the system can be represented graphically. Figure 1.4 plots the tax on banking income in percent as a function of the distortion (inefficiency) $\hat{\gamma} - \gamma^*$. The optimal tax is zero when there is no distortion, and increases with the size of the distortion as expected. The closed-form solution of the system defining the tax on banking income as a function of all exogenous parameters is:

$$\tau = \frac{2\delta(1 - \delta)\beta^e F_{HH}}{(2 - \beta^e)} \left[\frac{2(1 - \beta^i)}{\beta^e} \frac{F_{HH} - F_{HL}}{F_{HH}} + \frac{1 - \beta^e}{\beta^e} \left(1 - 2\delta \frac{F_{HH} - F_{LH}}{F_{HH}} - 1 - \frac{2\delta - \beta^e}{2\delta\beta^e(1 - \delta)} \right) \right].$$

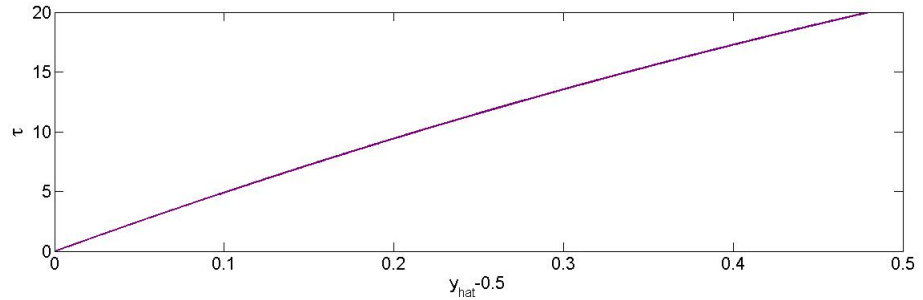


Figure 1.4: Tax on financial income vs. inefficiency

1.2.8 Comparative statics

Returning to the solution of the decentralized equilibrium, I analyze the comparative statics of the outcome of the model as exogenous parameters change.

The decentralized equilibrium is a function of all exogenous parameters: $\hat{\gamma} = f(\delta, \beta^e, z^H/z^L, \beta^i, k^H/k^L)$. For example, Figure 1.5 presents the solution $\hat{\gamma}$ as a function of the bargaining power δ . As we can see, the decentralized equilibrium exists only for $\delta \in [0, \hat{\delta}]$; there is no solution for $\delta > \hat{\delta}$. The decentralized equilibrium coincides with the constrained efficient outcome only for one particular value of the bargaining power $\hat{\delta}$.

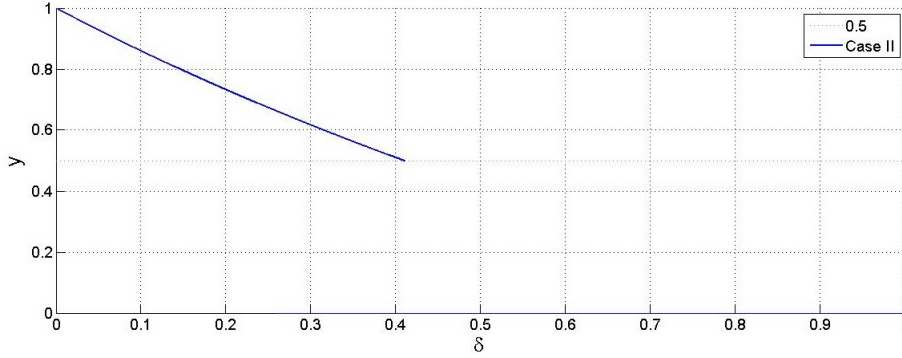


Figure 1.5: Fraction of bankers vs. bargaining power of entrepreneur (efficient fraction is 1/2)

Figure 1.6 presents the solution of the decentralized equilibrium as a function of wealth k^H/k^L and talent z^H/z^L dispersion. As we can see, wealth dispersion has a stronger impact on the size of the financial sector. More importantly, the static model predicts that an increase in wealth inequality will be associated with the growth of finance. When the rich get richer, they demand more finance. This is in line with empirical evidence. However, the wealth distribution has been considered completely exogenous up until now. The next section endogenizes the wealth distribution by introducing dynamics into the model.

1.3 Dynamic model and quantitative analysis

1.3.1 Dynamic model

As we saw above, the joint distribution of wealth and talent is an important determinant of the size of the financial sector and the degree of inefficiency. While

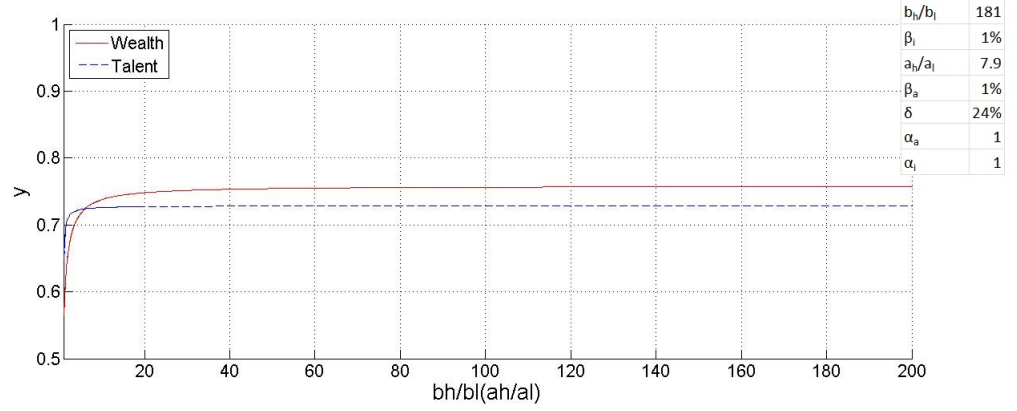


Figure 1.6: Fraction of bankers vs. dispersion of wealth (talent)
(efficient fraction is 1/2)

the distribution of talent is often considered exogenous, it is difficult to think about the wealth distribution as a fully exogenous object. In this subsection, I allow for endogenous wealth accumulation. Endogenous growth of wealth inequality leads to expansion of the financial sector. The rich get richer because they can afford to pay high fees for financial services, which yield a higher return on their savings. The higher are fees, the more talented agents work in finance. Consequently, the growth of finance and the increase in wealth inequality go hand in hand.

To introduce simple dynamics, I consider an infinite overlapping generations (OLG) model. The OLG structure seems to be natural for two reasons. First, I study relatively long-term growth of wealth and the size of finance (both have grown for at least the last six decades). Second, the generation structure is well suited to the problem, because agents undergo an interesting life cycle with low-wealth young age and higher-wealth old age. The young make an occupational choice and work in one of the two sectors. The middle-aged invest the wealth they have accumulated while young. The old consume the results of this investment.

I adopt the most basic OLG model. Every individual maximizes lifetime consumption and lives for three periods: youth, middle age and old age. Individuals are born in time t , work at time t , receive their income at $t + 1$ and consume at $t + 2$. Individuals pass through three stages over the life cycle: working, invest-

ment and consumption. The young are endowed with talent z and no wealth. The young make an occupational choice either to be an entrepreneur or a banker. The middle-aged are investors because they have wealth k , which they accumulated while young. The middle-aged have a choice of either being matched randomly or paying to a banker the price p_i in exchange for being matched with certainty with a talented entrepreneur. The middle-aged have no talent, because it fully depreciated within one period. The old consume the result of their investments.⁶

The rest remains as before. Individuals, who are born every period, can be of two types: talented or ordinary. Individuals are assumed to be risk-neutral and not to discount the future. The production function $F(z, k)$ is strictly supermodular and depends on both capital and talent. Financial frictions are two-sided private information and one-to-one matching. The high-type z^h banker can provide intermediated matching, while the low-type z^L banker can provide only random matching.

To keep two types of wealth, I consider a stand-in household that abstracts from the distinction between expected and realized income. Following Lucas and Rapping (1969) and more recent examples (Rogerson, 2008; Gertler and Kiyotaki, 2010), the stand-in household assumption has been a popular tool in macroeconomics to keep models tractable. I introduce the stand-in household in the following way. First, there is income sharing in finance. The realized income that every banker receives is the same as her expected income. Hence, all young talented agents z^H receive the same income, and become capital-abundant investors when they are old. Second, there is pooling of investment funds to ensure that the realized income that an ordinary entrepreneur receives is the same as her expected income. Hence, all ordinary entrepreneurs receive the same amount of capital.⁷

⁶Alternatively, due to risk-neutrality, individuals find it optimal to save their income fully and consume only in the last period. The age-related decline of cognitive abilities is a well-established fact in psychology. There is no consensus regarding the magnitude of the effect or the exact mechanism. The wealth-age profile is also well documented. Wealth grows rapidly over the life cycle and reaches its peak during one's 60s (the end of working age) and flattens or slightly declines afterwards.

⁷We can think of this as an insurance scheme within the financial sector. If agents are slightly

These assumptions change nothing from the point of view of expected incomes, but keep the model tractable. If I dropped any of these assumptions, the number of types would grow exponentially with a constant doubling every period.

The simple model produces life-cycle behavior consistent with the data: agents with a given talent level undergo a relatively realistic life cycle with low-income working youth, high-income investment middle age, and retirement with high consumption and zero income. Individuals typically start to accumulate assets for their retirement during middle age, around the age of 40 (Gourinchas and Parker, 2002). Wealth grows rapidly over the life cycle, reaches its peak at the age of 60 and flattens out afterwards. Total individual consumption, including housing and non-housing consumption, mimics individuals' wealth (Yang, 2009).

I keep the distribution of talent constant over time, and assume an initial distribution of wealth parametrized by the share of capital-abundant investors β_0^i and their wealth k_0^H , and the wealth of capital-scarce investors k_0^L . To use the solution of the static model from the previous subsection, I need to define the evolution of the wealth distribution. The system of equations below defines the evolution of the wealth distribution in the model:

$$\beta_t^i = \beta^e, \quad (1.15)$$

$$k_{t+1}^H = x_t^e = \delta \left(F(z^H, k_t^H) - d_t^i \right) + (1 - \delta) d_t^e, \quad (1.16)$$

$$k_t^M = \frac{k_t^H (\beta_t^i - \beta^e (1 - \hat{\gamma}_t)) + k_t^L * (1 - \beta_t^i)}{1 - \beta^e (1 - \hat{\gamma}_t)}, \quad (1.17)$$

$$k_{t+1}^L = \delta F(z^L, k_t^M). \quad (1.18)$$

Due to profit sharing, all talented agents receive the same income and become investors next period. Hence the share of capital-abundant investors every period, with the exception of the first one, is equal to $\beta_{t+1}^i = \beta^e$, expression (1.15). The next-period wealth of capital-abundant investors k_{t+1}^H is defined by expression (1.16) using the expressions for outside options in the case of intermediated matching (1.10) and (1.9). Finally, I define the next-period wealth of capital-scarce investors k_{t+1}^L , expression (1.18). Due to fund pooling, every entrepreneur

risk averse, $u_{t+1}^o = (c_{t+1}^o)^{1-\epsilon}$, where $\epsilon \approx 0$, all bankers are willing to engage in income sharing, and all investors are willing to engage in fund pooling.

who is not matched receives the same amount of funds k_t^M , expression (1.17), and consequently the same income which becomes her next-period wealth.

The next subsection brings the model to the US data in an attempt to replicate the dynamics of wealth and the financial sector.

1.3.2 The US experience

This theoretical model has been designed to explain how the role and size of the financial sector is determined and whether this size is efficient. Even though the model is simplistic, the calibrated version of it performs surprisingly well. The goal of the dynamic model is to explain the interrelationship between the growth of the financial sector in terms of employment and the growth of wealth. Therefore, I choose them as data moments to be matched.

In the first calibration exercise, I seek to explain the behavior of the whole financial sector. Then, I recalibrate the model to explain the behavior of one subindustry of the financial sector—*private equity finance*. Even though the model can be applied to the financial sector as a whole, private equity finance is a subindustry for which the assumptions of the model are particularly valid: matching and information superiority. In particular, a private equity fund precisely does matching between a few selected startups and high-net-worth individuals. The private equity fund provides an opportunity to invest in a few companies over a long-term horizon for a small number of wealthy investors.

The first eight parameters described in Table 1.2 are used to match as closely as possible the share of employment in finance and the ratio of top 5% wealth to median wealth over time in the US. The economy starts initially with an almost egalitarian distribution of wealth ($k^H = 100$ vs. $k^L = 95$); otherwise the share of employment in finance immediately jumps to the steady-state level and the wealth disparity explodes. The distribution of talent remains the same every period: talented agents are assumed to be 1.7 ($z^H/z^L = 6.5/3.8 = 1.7$) times more talented than ordinary ones. Following Romer (1986), the production function exhibits non-decreasing return to scale with respect to capital ($\alpha_k = 1.095$, $\alpha_z = 1$)—it is very similar to the AK production function. While the increasing return on capital

Table 1.2: Parameter values

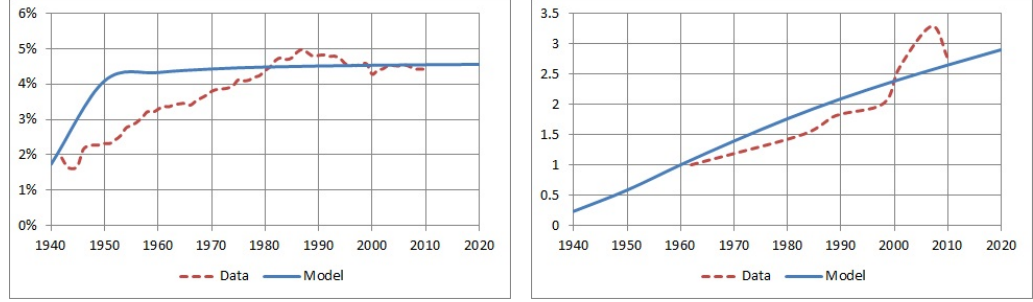
Parameter		Value
Distribution of talent		
Talented	z^H	6.5
Ordinary	z^L	3.8
Share of talented	β^e	6.7%
Initial distribution of wealth		
Capital-abundant	k_0^H	100
Capital-scarce	k_0^L	95
Share of capital-abundant	β_0^i	3%
Other parameters		
Elasticity of talent	α_z	1
Elasticity of capital	α_k	1.095
Bargaining power of entrepreneur	δ	21%

generates the growth of aggregate capital, the talent differentials ensure the rise of wealth dispersion. Choosing realistic values for the eight parameters, I recall the definition of $\hat{\delta}$, the maximum entrepreneurial bargaining power consistent with the existence of an equilibrium $\hat{\gamma}$, as a function of other parameters from Proposition 2. (See appendix 1.6.2 for more detail.) The calculated value is $\hat{\delta}_0 = 36.6\%$, and it is growing with wealth dispersion, while the data suggests that the level of entrepreneurial bargaining power is rather small. (Kaplan and Stromberg (2003) report that the average founders' share equals 21.3% of a portfolio company's equity value.) Hence, I set the level of bargaining power to be 21% and keep it constant over time. Since $\delta < \hat{\delta}_t$, according to Proposition 2, the solution of the decentralized equilibrium exists and is inefficient ($\hat{\gamma}_t > \gamma^*$). The inefficiency is growing over time because of increasing wealth dispersion. Table 1.2 summarizes all parameter values.

Figure 1.7 shows the comparison between the data and the outcome of the model. On the left-hand side, we can see the share of employment in finance over time. On the right-hand side is the ratio of top 5% wealth to median wealth over time. As we can see, while the share of employment in finance was growing until the 1980s and then stabilized a little above 5%, the ratio of top 5% wealth to median wealth has been increasing over the whole sample apart from a small drop during the Great Recession. The drop reflects the sharp decrease of asset prices: stocks, housing etc. This business cycle type of consideration is outside the scope of this paper.

The length of the period is a decade. A decade is arguably the shortest scale to study long-term events, such as the accumulation of wealth and structural changes in employment, and possibly the most appropriate one. First, ten years is a standard investment horizon for long-term investors, who are the subjects of this study. Second, a decade is a long enough period to abstract away from business cycle considerations, which are outside the scope of this paper. Furthermore, most of the data is available only for the last 60 years.

To test the external validity of the model, let us consider some other interesting data trends. First, the share of finance in GDP has steadily increased over the



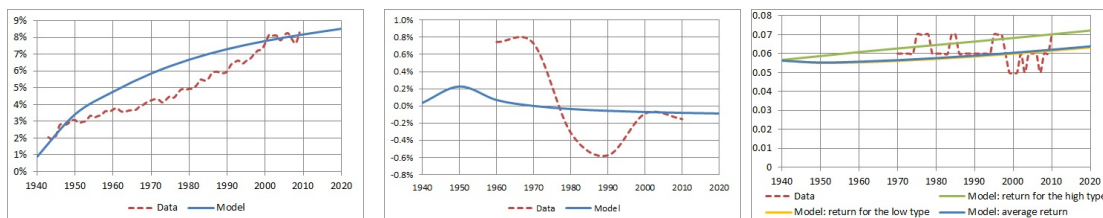
(a) Share of employment in finance

(b) Top 5% wealth to median wealth

Figure 1.7: US data moments

whole period despite the stabilization of the employment share. Figure 1.8a shows that this is in line with my model. Second, Figure 1.8b represents the deviation ten-year moving average growth rate of US productivity from the long-term average growth rate for the whole period of observation. It is calculated based on labor productivity per hour worked in 1990 US dollars (converted at Geary–Khamis PPPs). The yearly average for the period after the Second World War is 1.9%. Figure 1.8b shows that US productivity growth has slowed significantly since 1973, with a minor resurgence of productivity in the 1990s due to the IT revolution. Several explanations of the slowdown have been suggested, but none has been found to be fully satisfactory. My model links the slowdown to a misallocation of talent. As we can see from the data and the model, the biggest increase of the employment share in finance happened in the 1980s. We observe the largest drop in productivity in the US in exactly the same period. This argument is in line with Nordhaus (1982), who argues that a depletion of investment opportunities due to a lack of inventions caused the slowdown. The patent data provides indirect support for this statement. Patent applications and the quality of patents declined significantly in the 1980s despite a constantly increasing number of scientists and researchers. We observe a substantial decline in the share of the US in world patents not only due to the rise of China, India and Korea, but also in comparison with Scandinavian countries (OECD, 2011). The Scandinavian countries have not experienced any substantial growth of the financial sector. To sum up, I argue

that the US productivity slowdown was partly caused by the decreasing number of talented individuals in industry to produce innovations.



(a) Share of finance in GDP (b) Productivity slowdown (c) Average return on wealth

Figure 1.8: Additional facts about the US

Third, Figure 1.8c shows the average return on private wealth in the US for the period 1970–2010. The return on wealth (capital) has oscillated around a central value of 5–7% a year. Piketty and Zucman (2014) first noticed that the rate of return on capital is greater than the rate of economic growth over the long term; the result is a growing concentration of wealth. The substantial dispersion in wealth accumulation has been observed despite the minor difference in returns on private wealth between capital-abundant and capital-scarce investors. Given that I do not try to match the average return, it is surprising that the chosen 10-year horizon generates a similar return on capital as in the data.

1.4 Cross-country data

In this section I provide cross-country evidence to answer the question of how the distribution of wealth and talent affects the size of the financial sector. As predicted by my model, the evidence clearly shows a positive relationship between the size of the financial sector and inequalities of wealth and talent. Even though my model predicts a causal link from the joint distribution of wealth and talent to the equilibrium size of the financial sector, in this section I intend to make no causal statement, only to document correlations.

To test the relationship, I need to have a compatible cross-country measure of moments of talent and wealth distributions, and the size of the financial sector.

Unfortunately, data availability limits my choice. For the talent distribution, I employ the scores in the PISA test. The PISA test aims to evaluate education systems worldwide every three years by assessing 15-year-olds' competencies in key subjects: reading, mathematics and science. To date, over 70 countries have participated in PISA. It is a widely used measure for cross-country comparisons of students' performance. Furthermore, it is highly unlikely that this measure of talent suffers from reverse causality. There is no reason why the size of finance today might affect the performance of high-school students today. I used the mean and variance of scores in the PISA test for the years 2003, 2006 and 2009. I choose the mean and variance of 2009 science scores in the PISA test as a proxy for the moments of talent distributions, because it includes the greatest number of countries. Moreover, this choice hardly affects the results, because PISA scores are highly correlated over time and disciplines: the correlation coefficients exceed 0.97.

To the best of my knowledge, there is no cross-country data on wealth inequality. Therefore, I have to use the income distribution as a proxy for the wealth distribution. Income inequality is a fairly standard proxy for wealth inequality, but possibly underestimates wealth inequality. Income and wealth are not particularly well correlated either at the individual level for a given point (Rodriguez et al. (2002) estimate the correlation between wealth and labor income to be 0.27) or across countries (Fredriksen, 2012). However, if we measure the correlation over time between top income and wealth shares for a particular country, for example the US, we observe that the shares are highly correlated. The more concentrated are the shares, the higher are the correlations between them. In addition, the income shares are more volatile and tend to lead the wealth shares. We can see from Figure 1.9 that the dynamics of wealth shares closely track the dynamics of income shares for the US.

To measure income inequality, I employ the Gini indexes from the Standardized World Income Inequality Database (SWIID) and top income shares from the World Top Incomes Database. The SWIID provides comparable Gini indexes of gross and net income inequality for 173 countries for as many years as possible from 1960.

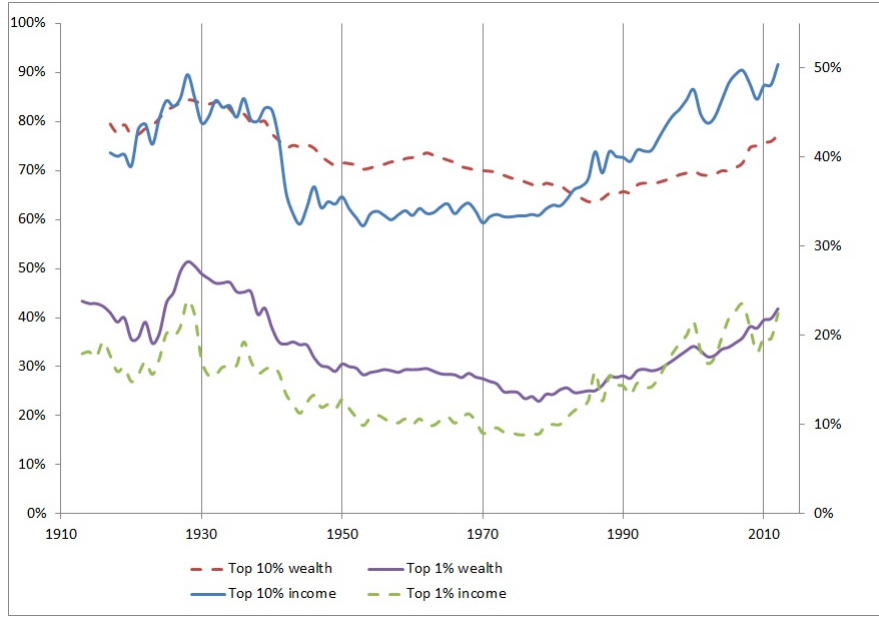


Figure 1.9: US top income and wealth shares

The correlation for 10% (1%) is 0.52 (0.77)

The World Top Incomes Database includes 45 countries for over a century for some countries.

The last issue is how to measure the size of finance. I construct the share of financial industry employment in total employment using two datasets: the International Labour Organization (ILO) dataset, which contains employment by economic activity for 165 countries starting from 1968, and the STAN Database for Structural Analysis, which contains industry-level data for employment and output for 15 OECD countries from the 1970s up to the present.

After conducting panel unit root tests, such as the Fisher combination test (Maddala and Wu, 1999) and the Pesaran (2007) panel unit root test, I conclude that real GDP per capita is non-stationary; therefore I compute its growth rate. The specification of the full model for the OLS estimation is given by

$$Emp_{it} = \gamma_0 + \gamma_1 gGDP_{it} + \gamma_2 II_{it} + \gamma_3 MP_{it} + \gamma_4 VP_{it} + \varepsilon_{it}, \quad (1.19)$$

where Emp_{it} is the share of employment in the financial sector; $gGDP_{it}$ is the growth rate of real GDP per capita; II_{it} is the measure of income inequality (Gini

or top shares); MP_{it} is the mean PISA score; and VP_{it} is the variance of the PISA score.

Table 1.3: The share of finance (in % of total employment)

	(1)	(2)	(3)	(4)
Top 5% income share	0.00460***	0.00404***		
Top 1% income share			0.00466***	
Gini index				0.000613*
GDPPC growth	0.0139	0.00693	-0.0239	0.00738
Variance of PISA score		0.00189*	0.00195***	0.00284***
Mean of PISA score		0.000284	0.000556***	0.000370***
Obs.	215	201	201	744
R^2	0.176	0.217	0.273	0.529

* (p<0.05), ** (p<0.01), *** (p<0.001)

The results from Table 1.3 are consistent with my model. A higher share of financial employment is associated with more unequal income and talent distributions. The estimation results from the biggest sample, the last column with the Gini coefficient, show this clearly. Income inequality has an even stronger effect if I use the top income share instead of the Gini coefficient (columns 2 and 3). The result is not driven by country fixed effects (FE). We can see by comparing column 1 with FE and column 2 that the estimated coefficient of the top 5% share remains positive, significant and almost unchanged.

1.5 Discussion and conclusion

In this section, I review the impact of the different assumptions on the outcome of the model: the inefficiency result and the inequality result. First, the inefficiency result states that the decentralized equilibrium is generically inefficient. Second, the inequality result states that the endogenous growth of wealth inequality leads to the expansion of the financial sector.

Preferences and technology: First, for simplicity I assume the Cobb–Douglas production function $z^{\alpha_z} k^{\alpha_k}$, which satisfies the supermodularity condition. However, the choice of a production function should not affect the results, because as long as z and k are not fully substitutable, the supermodularity condition holds. (See Topkis (1998) for a comprehensive mathematical treatment of supermodularity, and Milgrom and Roberts (1990) and Vives (2005) for applications in game theory and economics.) According to the Becker–Brock theorem, the supermodularity condition implies that positive assortative matching is the first best allocation of my model. Hence, the results of the model remains unchanged as long as the production function is supermodular. Second, it should not be complicated to include labor as an additional input, but it would not add further insights on the questions addressed in this paper. It should not affect the choice of talented agents, but it might have interesting implications for ordinary agents. Third, if we consider a risk averse utility function instead of a risk neutral one, all agents would like to engage in risk sharing. If profit sharing and fund pooling are available options, the introduction of risk aversion does not change anything, because expected and realized incomes are the same. If these options are not available, the impact of risk aversion is ambiguous. On the one hand, investors are willing to pay a higher price for intermediated matching. The higher is the price, the higher is the income of a banker. On the other hand, due to higher uncertainty with respect to this income, risk aversion makes a banking career a less attractive option.

Distribution of types: First, as long as within each period the wealth distribution is independent from the talent distribution, the investment decision is independent from the occupational choice. This makes the solution of the problem tractable. The consideration of the two-dimensional joint distribution of wealth and talent complicates the analysis enormously without much additional insight for this particular question. Second, the fact that the constrained efficient allocation admits only two values is an artifact of the discrete distribution of talent and the particular type of information advantage for talented agents in finance: a banker with ability z can distinguish between ideas with productivity z and $z' < z$. As long as both assumptions hold, the constrained efficient allocation admits two

values (zero and one-half) for each type of talent: the planner would find it optimal either to keep the allocation under random matching or make it as close as possible to the allocation under assortative matching by exhausting fully the opportunities for intermediated matching. The allocation in the case of a continuous talent distribution would strongly depend on the assumption made with respect to the impact of talent on agents' productivity in the two sectors. Third, I can provide intuition for the case of a continuous wealth distribution. The constrained efficient solution either have a positive share for all values of talent distribution z or there exists a threshold in terms of ability \bar{z} , that separates bankers and entrepreneurs. Calculating the decentralized equilibrium is a complicated numerical task.

Different types of frictions: First, this paper focuses on how the financial sector arises as a result of one type of relevant friction, adverse selection. The financial sector clearly provides other useful functions to the economy: it allocates not only information, but also decision power and risk. On the theoretical side, the financial sector's functions include: screening to mitigate the effect of adverse selection; monitoring to prevent the effects of moral hazard; auditing and punishment to mitigate the effects of opportunistic behavior in the context of costly state verification. (See Freixas and Rochet (2008) for a precise survey of the theoretical literature.) Comparing my model with that of Bolton et al. (2011), who study moral hazard and the financial sector as a liquidity provider, while I consider adverse selection and the financial sector as a classical intermediary, we both obtain a similar result in terms of efficiency, but the mechanisms are substantially different. This suggests that the misallocation result might be a general feature of models with financial frictions. Under the assumption that talent in finance affects the efficiency of monitoring, the inequality result is likely to survive as well. Second, The matching friction is clearly important for the inefficiency, because in the perfectly competitive market prices would take into account the negative externality, which arrived from the occupational choice. However, More a general form of matching friction, many-to-one matching can be easily introduced into the environment in at least two ways: through diminishing returns on capital and fixed costs of engaging with investors; or through making entrepreneurs' bargaining power depend posi-

tively on the number of investors in the market. *Ceteris paribus*, it is likely that many-to-one matching would lead to more inefficiency in comparison to one-one matching. The more investors can be matched with one entrepreneur, the fewer bankers are needed to restore efficiency. The income of a banker increases with the number investors matched with one entrepreneur. Hence, an even larger fraction of talented agents is attracted to finance. The inefficiency should increase due to both a decline in the constrained efficient fraction of talented agents in finance and a rise in the decentralized one.

Third, the issue of competition has been studied extensively. Monopoly is usually viewed as a bad thing. However, in my framework one monopolistic firm in the financial sector might restore efficiency, because it maximizes the total surplus by pushing all agents to their outside option. The monopolist is always on the short side of the market. It would set the prices for its services to make both entrepreneurs and investors indifferent between paying for the services and being matched, and being randomly matched for free. On top of this, the monopolist can set wages for its workers (bankers) to make them indifferent between the two sectors. Hence, the monopolist could extract the total surplus and would hire the efficient number of bankers. However, this possible advantage of a monopoly in the context of information provision does not overcome common disadvantages of monopoly for a society.

1.5.1 Conclusion

This paper develops a new model of an economy with a financial sector and heterogeneous agents. The model sheds light on the role of the financial sector and its implication for the allocation of capital between entrepreneurs and the allocation of talent between finance and industry. Talent is important for both industry and the financial sector: more talent in industry means more output is produced, while more talent in finance means capital is allocated more efficiently. The model establishes a link between the growth of the financial sector and the increase in wealth inequality. It shows that the market overproduces finance, but this inefficiency can be corrected by taxing bankers' income. For the future, it

would be interesting to quantitatively assess the size of this inefficiency and the importance of current wealth in limiting investment.

Bibliography

- BAUER, C. AND J. V. R. MORA (2014): “The Joint Determination of TFP and Financial Sector Size,” Working paper.
- BECKER, G. S. (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81, 813–46.
- BERKES, E., U. PANIZZA, AND J.-L. ARCAND (2012): “Too Much Finance?” IMF Working Papers 12/161, International Monetary Fund.
- BIENZ, C. AND J. HIRSCH (2011): “The Dynamics of Venture Capital Contracts,” *Review of Finance*, 16, 157–195.
- BOLTON, P., T. SANTOS, AND J. A. SCHEINKMAN (2011): “Cream skinning in financial markets,” NBER Working Papers 16804, National Bureau of Economic Research, Inc.
- BRUNNERMEIER, M. K., T. M. EISENBACH, AND Y. SANNIKOV (2012): “Macroeconomics with Financial Frictions: A Survey,” NBER Working Papers 18102, National Bureau of Economic Research, Inc.
- BUERA, F. J. AND Y. SHIN (2013): “Financial Frictions and the Persistence of History: A Quantitative Exploration,” *Journal of Political Economy*, 121, 221 – 272.
- CAHUC, P. AND E. CHALLE (2012): “Produce Or Speculate? Asset Bubbles, Occupational Choice, And Efficiency,” *International Economic Review*, 53, 1105–1131.

- CHEN, Q., I. GOLDSTEIN, AND W. JIANG (2006): “Price Informativeness and Investment Sensitivity to Stock Price,” *Review of Financial Studies*, 20, 619–650.
- COOLEY, T. F., R. MARIMON, AND V. QUADRINI (2013): “Risky Investments with Limited Commitment,” NBER Working Papers 19594, National Bureau of Economic Research, Inc.
- DURNEV, A., R. MORCK, AND B. YEUNG (2004): “Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation,” *Journal of Finance*, 59, 65–105.
- FREDRIKSEN, K. B. (2012): “Less Income Inequality and More Growth - Are they Compatible? Part 6. The Distribution of Wealth,” OECD Economics Department Working Papers No. 929, OECD.
- FREIXAS, X. AND J.-C. ROCHET (2008): *Microeconomics of Banking, 2nd Edition*, MIT Press Books, The MIT Press.
- GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2013): “Finance and the Preservation of Wealth,” NBER Working Papers 19117, National Bureau of Economic Research, Inc.
- GERTLER, M. AND N. KİYOTAKI (2010): “Financial Intermediation and Credit Policy in Business Cycle Analysis,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman and M. Woodford, Elsevier, vol. 3 of *Handbook of Monetary Economics*, chap. 11, 547–599.
- GOLDIN, C. AND L. F. KATZ (2008): “Transitions: Career and Family Life Cycles of the Educational Elite,” *American Economic Review*, 98, 363–69.
- GOLDSMITH, R. (1995): *Financial Structure and Development*, vol. 1, Yale University Press.
- GORDON, P. (2000): “The Americas,” *Global Finance*, 14.
- GOURINCHAS, P.-O. AND J. A. PARKER (2002): “Consumption Over the Life Cycle,” *Econometrica*, 70, 47–89.

- GREENSPAN, A. (2002): “World Finance and Risk Management,” <http://www.federalreserve.gov/BoardDocs/Speeches/2002/200209253/default.htm>.
- GREENWOOD, R. AND D. SCHARFSTEIN (2013): “The Growth of Finance,” *Journal of Economic Perspectives*, 27, 3–28.
- HOSIOS, A. J. (1990): “On the Efficiency of Matching and Related Models of Search and Unemployment,” *Review of Economic Studies*, 57, 279–98.
- HSIEH, C.-T. AND P. J. KLENOW (2009): “Misallocation and Manufacturing TFP in China and India,” *The Quarterly Journal of Economics*, 124, 1403–1448.
- JOVANOVIC, B. (2014): “Misallocation and Growth,” *American Economic Review*, 104, 1149–71.
- JOVANOVIC, B. AND B. SZENTES (2013): “On the Market for Venture Capital,” *Journal of Political Economy*, 121, 493 – 527.
- KAPLAN, S. N. AND P. STROMBERG (2003): “Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts,” *Review of Economic Studies*, 70, 281–315.
- KNEER, C. (2012): “The Absorption of Talent into Finance: Evidence from U.S. Banking Deregulation,” *Job Market Paper of Tilburg University*, 1–28.
- LEVINE, R. (2005): “Finance and Growth: Theory and Evidence,” in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1 of *Handbook of Economic Growth*, chap. 12, 865–934.
- LEVINSOHN, J. AND A. PETRIN (2003): “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 70, 317–341.
- LUCAS, ROBERT E, J. AND L. A. RAPPING (1969): “Real Wages, Employment, and Inflation,” *Journal of Political Economy*, 77, 721–54.
- LUCAS, R. J. (1988): “On the mechanics of economic development,” *Journal of Monetary Economics*, 22, 3–42.

- LUO, Y. (2005): “Do Insiders Learn from Outsiders? Evidence from Mergers and Acquisitions,” *The Journal of Finance*, 60, 1951–1982.
- MADDALA, G. S. AND S. WU (1999): “A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test,” *Oxford Bulletin of Economics and Statistics*, 61, 631–52.
- MAYER-HAUG, K., S. READ, J. BRINCKMANN, N. DEW, AND D. GRICHNIK (2013): “Entrepreneurial talent and venture performance: A meta-analytic investigation of SMEs,” *Research Policy*, 1–23.
- MERTON, R. C. (1995): *The Global Financial System: A Functioning Perspective*, Harvard Business School Press.
- MGI (2011): “Big data: The next frontier for innovation, competition, and productivity,” The mckinsey global institute report, The McKinsey Global Institute.
- MICHELACCI, C. AND J. SUAREZ (2004): “Business Creation and the Stock Market,” *Review of Economic Studies*, 71, 459–481.
- MIDRIGAN, V. AND D. Y. XU (2014): “Finance and Misallocation: Evidence from Plant-Level Data,” *American Economic Review*, 104, 422–58.
- MILGROM, P. AND J. ROBERTS (1990): “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,” *Econometrica*, 58, 1255–77.
- MURPHY, K. M., A. SHLEIFER, AND R. W. VISHNY (1991): “The Allocation of Talent : Implications for Growth,” *The Quarterly Journal of Economics*, 106, 503–530.
- NASH, J. (1950): “The Bargaining Problem,” *Econometrica*, 18, 155–162.
- (1953): “Two-Person Cooperative Games,” *Econometrica*, 21, 128–140.
- NORDHAUS, W. D. (1982): “Economic policy in the face of declining productivity growth,” *European Economic Review*, 18, 131–157.

- OBAMA, B. (2009): "The President Explains His Larger Vision on the Economy," <http://www.whitehouse.gov/video/The-President-Explains-His-Larger-Vision-on-the-Economy>.
- OECD (2011): "OECD Science, Technology and Industry Scoreboard," Oecd report, OECD.
- PESARAN, M. H. (2007): "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 22, 265–312.
- PHILIPPON, T. (2010): "Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?" *American Economic Journal: Macroeconomics*, 2, 158–182.
- PHILIPPON, T. AND A. RESHEF (2012): "Wages and Human Capital in the U.S. Finance Industry: 1909–2006," *The Quarterly Journal of Economics*, 127, 1551–1609.
- PIKETTY, T. AND G. ZUCMAN (2014): "Capital is Back: Wealth-Income Ratios in Rich Countries, 1700–2010," *The Quarterly Journal of Economics*.
- RESTUCCIA, D. AND R. ROGERSON (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Plants," *Review of Economic Dynamics*, 11, 707–720.
- RODRIGUEZ, S. B., J. DIAZ-GIMENEZ, V. QUADRINI, AND J.-V. RIOS-RULL (2002): "Updated facts on the U.S. distributions of earnings, income, and wealth," *Quarterly Review*, 2–35.
- ROGERSON, R. (2008): "Structural Transformation and the Deterioration of European Labor Market Outcomes," *Journal of Political Economy*, 116, 235–259.
- ROMER, P. M. (1986): "Increasing Returns and Long-Run Growth," *The Journal of Political economy*, 94, 1002–1037.
- SAHLMAN, W. A. (1990): "The structure and governance of venture-capital organizations," *Journal of Financial Economics*, 27, 473–521.

- SANTARELLI, E., L. KLUMP, AND A. THURIK (2006): “Gibrat’s Law: An Overview of the Empirical Literature,” in *Entrepreneurship, Growth, and Innovation*, ed. by E. Santarelli, Springer US, vol. 12 of *International Studies in Entrepreneurship*, 41–73.
- SCHUMPETER, J. A. (1934): *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, Transaction Publishers.
- SHU, P. (2012): “Essays on Innovation, Productivity, and Talent Allocation,” *MIT Job Market Paper*.
- TOBIN, J. (1984): “On the efficiency of the financial system,” *Lloyds Bank Review*, 153, 1–15.
- TOPKIS, D. M. (1998): *Supermodularity and Complementarity*, Princeton University Press, first edition ed.
- VIVES, X. (2005): “Complementarities and Games: New Developments,” *Journal of Economic Literature*, 43, 437–479.
- VOLCKER, P. (2009): “Paul Volcker: Think More Boldly,” <http://online.wsj.com/articles/SB10001424052748704825504574586330960597134>.
- WADHWA, V., B. A. RISSING, AND G. GEREFFI (2006): “Industry Trends in Engineering Offshoring,” report, Duke University.
- YANG, F. (2009): “Consumption over the Life Cycle: How Different is Housing?” *Review of Economic Dynamics*, 12, 423–443.

1.6 Appendix A

1.6.1 Proof of Proposition 1

The proof is organized in the following way. First, I show that as long as γ is greater than $1/2$, aggregate output $Y(\gamma)$ decreases with γ . Second, depending on other parameters, $Y(\gamma)$ is either a strictly increasing, or strictly decreasing, or a non-monotonic function of γ for the whole interval $\gamma \in [0, 1/2]$. In the latter case, I prove that the function is a convex function for $\gamma \in [0, 1/2]$.

For $\gamma > 1/2$, aggregate output $Y(\gamma)$ is given by

$$Y(\gamma) = \frac{(\beta^i - (1-\gamma)\beta^e)}{(1-(1-\gamma)\beta^e)}(1 - \beta^e)F_{LH} + \frac{(1-\beta^i)}{1-(1-\gamma)\beta^e}(1 - \beta^e)F_{LL} + (1 - \gamma)\beta^e F_{HH}. \quad (1.20)$$

To show whether it is increasing or decreasing, we take the derivative of (1.20) with respect to γ :

$$\frac{\partial Y}{\partial \gamma} \Big|_{\gamma > 1/2} = \frac{\beta^e(1-\beta^i)(1-\beta^e)}{(1-(1-\gamma)\beta^e)^2} F_{LH} - \beta^e F_{HH} - \frac{\beta^e(1-\beta^i)}{(1-(1-\gamma)\beta^e)^2} F_{LL}. \quad (1.21)$$

We need to estimate the sign of expression (1.21). Since F is a supermodular function, $F_{LL} \geq 0$ and $F_{HH} \geq F_{LH}$. Hence

$$\frac{\partial Y}{\partial \gamma} \Big|_{\gamma > 1/2} \leq \frac{\beta^e(1-\beta^i)(1-\beta^e)}{(1-(1-\gamma)\beta^e)^2} - \beta^e. \quad (1.22)$$

Expressing the two terms on the right-hand side of condition (1.22) using a common denominator, then replacing γ in the numerator by $1/2$ (this value makes the numerator as small as possible), and finally expanding, we obtain:

$$\frac{\partial Y}{\partial \gamma} \Big|_{\gamma > 1/2} \leq \beta^e \frac{(1-\beta^e)(1-\beta^i) - (1-\beta^e + \gamma\beta^e)^2}{(1-(1-\gamma)\beta^e)^2} \leq \beta^e \frac{(1-\beta^e)(1-\beta^i) - (1-\beta^e + 0.25(\beta^e)^2)}{(1-(1-\gamma)\beta^e)^2} \leq 0. \quad (1.23)$$

Here is the end of the proof of the first part. To prove the second part, I follow a similar procedure. I calculate the first derivative and restrict my attention to the case in which the first derivative is neither positive or negative for the whole interval $\gamma \in [0, 1/2]$. Then, I show that in this case the second derivative is positive, i.e. the function is convex.

For $\gamma \in [0, 1/2]$, the aggregate output $Y(\gamma)$ is given by

$$Y(\gamma) = \frac{(\beta^i - \gamma\beta^e)}{(1 - \gamma\beta^e)} [(1 - 2\gamma)\beta^e F_{HH} + (1 - \beta^e)F_{LH}] + \frac{(1 - \beta^i)}{(1 - \gamma\beta^e)} [(1 - 2\gamma)\beta^e F_{HL} + (1 - \beta^e)F_{LL}] + \gamma\beta^e F_{HH}. \quad (1.24)$$

Calculating the first derivative from (1.24), we obtain

$$\frac{\partial Y}{\partial \gamma} \Big|_{\gamma \in [0, 1/2]} = \frac{1}{(1 - \gamma\beta^e)^2} [\beta^e F_{HH} (1 + 2\gamma\beta^e - (\gamma\beta^e)^2 + \beta^i\beta^e - \beta^e + 2\beta^i) - (1 - \beta^e)\beta^e(1 - \beta^i)(F_{LH} - F_{LL}) - (2 - \beta^e)\beta^e(1 - \beta^i)F_{HL}]. \quad (1.25)$$

The first derivative is negative for the whole interval $\gamma \in [0, 1/2]$ if

$$\beta^e F_{HH} (1 + 2\gamma\beta^e - (\gamma\beta^e)^2 + \beta^i\beta^e - \beta^e + 2\beta^i) - (2 - \beta^e)\beta^e(1 - \beta^i)F_{HL} < (1 - \beta^e)\beta^e(1 - \beta^i)(F_{LH} - F_{LL}). \quad (1.26)$$

The left-hand side of inequality (1.26) increases with γ , while the right-hand side of inequality (1.26) is independent of γ . Hence if inequality (1.26) holds for $\gamma = 1/2$, it holds for any $\gamma \in [0, 1/2]$.

$$[\beta^e F_{HH} (1 - 0.25(\beta^e)^2 + \beta^i\beta^e + 2\beta^i) - (1 - \beta^e)\beta^e(1 - \beta^i)(F_{LH} - F_{LL}) - (2 - \beta^e)\beta^e(1 - \beta^i)F_{HL}] < 0 \quad (1.27)$$

If inequality (1.27) holds, the first derivative is negative. In the opposite case, the sign of the derivative is unknown. Inequality (1.27) imposes the restriction on the set of exogenous parameters.

We now calculate the second derivative and check its sign:

$$\frac{\partial^2 Y}{\partial \gamma^2} \Big|_{\gamma \in [0, 1/2]} = \frac{2\beta^e}{(1 - \gamma\beta^e)^3} [\beta^e F_{HH} (1 + \beta^i\beta^e + 2\beta^i) - (1 - \beta^e)\beta^e(1 - \beta^i)(F_{LH} - F_{LL}) - (2 - \beta^e)\beta^e(1 - \beta^i)F_{HL}]. \quad (1.28)$$

If the second derivative is positive, the function is convex. I need to show that the right-hand side of (1.28) is positive. As we have seen, when inequality (1.27) does not hold, the sign of the first derivative is unknown, but it imposes the restriction on the set of exogenous parameters. This is the case in which we need to know the sign of the second derivative. We can estimate the right-hand side using the complementary inequality to condition (1.27):

$$\beta^e F_{HH} (1 + \beta^i\beta^e + 2\beta^i) - (2 - \beta^e)\beta^e(1 - \beta^i)F_{HL} - (1 - \beta^e)\beta^e(1 - \beta^i)(F_{LH} - F_{LL}) \geq \beta^e F_{HH} 0.25(\beta^e)^2 > 0. \quad (1.29)$$

This completes the proof of the first part. We show that the sign of the first derivative is either negative or unknown. In the case in which it is unknown, we prove that the second derivative is strictly positive. Hence, the solution of the planner's problem can be either 0 or $1/2$. \square

1.6.2 Proof of Proposition 2

The characterization of a decentralized equilibrium is the following triplet: two prices and the share of talented agents in finance (p^i, p^e, γ) . We have:

$$\begin{aligned} p^i &\leq (1 - \delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e &\leq \delta(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ \frac{\min\{\gamma, 1-\gamma\}}{\gamma}(p^i + p^e) &= \delta(F_{HH} - d^i(\gamma)) + (1 - \delta)d^e(\gamma) - p^e. \end{aligned} \tag{1.30}$$

As a reminder, there are three types of agents who affect the solution of intermediated matching: capital-abundant investors, talented entrepreneurs and bankers. The number of investors is β^i , the equilibrium number of bankers is $\gamma\beta^e$, and the equilibrium number of entrepreneurs is $(1 - \gamma)\beta^e$. There are two markets and consequently two prices that clear them: entrepreneur–banker and investor–banker. System (1.30) should be solved differently depending on who is on the short side of both markets: capital-abundant investors, talented entrepreneurs or bankers. I show that a solution exists only if capital-abundant investors are on the long side of the investor–banker market. Furthermore, the condition $\gamma^* = 1/2$ imposes an additional restriction on the set of exogenous parameters, and eliminates a possible solution with bankers being on the short side with respect to talented entrepreneurs on the entrepreneur–banker market. However, the solution does not always exist. I state the existence condition as well.

The system can be solved backwards. First, we need to define who is on the short side of the market: capital-abundant investors, talented entrepreneurs or bankers. Second, I solve the random matching problem for a given size of the financial sector $\gamma\beta^e$ to determine the outside options of capital-abundant investors $d^i(\gamma)$ and talented entrepreneurs $d^e(\gamma)$ in the case in which they decide not to be matched with a high-type counterpart with certainty through a banker. Third,

using the solution of random matching as outside options, I solve the intermediated matching problem for capital-abundant investors and talented entrepreneurs.

Capital-abundant investors are on the short side: The number of capital-abundant investors is lower than the number of bankers who provide services for investors $\beta^i < \gamma\beta^e$. Hence, competition among bankers drives the price p^i down to zero. If the number of bankers is greater than the number of talented entrepreneurs, the bankers' income is zero, because prices for their service are zero.

Furthermore, the number of bankers cannot be greater than the number of talented entrepreneurs. Otherwise, bankers' income is zero, and any talented agent strictly prefers to be an entrepreneur. Thus, if capital-abundant investors are on the short side of the investor–banker market, the share of talented agents in finance must be $\gamma \leq 1/2$.

If $\gamma \leq 1/2$ and $p^i = 0$, the system (1.30) collapses to one condition:

$$\begin{aligned} p^e &= \delta(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ 2p^e &= \delta(F_{HH} - d^i(\gamma)) + (1 - \delta)d^e(\gamma). \end{aligned} \tag{1.31}$$

Substituting prices, system (1.31) collapses to condition (1.32)

$$\delta F_{HH} = \delta d^i(\gamma) + (1 + \delta)d^e(\gamma). \tag{1.32}$$

Condition (1.32) does not hold unless $\delta = 0$. Hence, capital-abundant investors cannot be on the short side in equilibrium.

Capital-abundant investors are on the long side: The number of capital-abundant investors is higher than the number of bankers who provide services for investors $\beta^i \geq \gamma\beta^e$. Hence, bankers push capital-abundant investors to their outside options. The first equation of system (1.30) becomes an equality. Two cases are possible.

First, if the number of bankers is lower than the number the talented entrepreneurs $\gamma \leq 1/2$, $1 - 2\gamma\beta^e$ talented entrepreneurs are left for random matching. We assume that investors have no access to a storage technology, while entrepreneurs have no opportunity for outside borrowing. Thus, the outside options for a random match—the set D in (1.5)—are $(0, 0)$. The solution of the Nash

bargaining problem gives the value of random matching for capital-abundant investors, which is equal to the probability of matching with somebody $\frac{1-2\gamma\beta^e}{1-\gamma\beta^e}$ multiplied by the sum of two terms: the probability of matching with a talented entrepreneur $\frac{(1-2\gamma)\beta^e}{1-2\gamma\beta^e}$ multiplied by the fraction of the project's output received by the investor $(1-\delta)F_{HH}$; and the probability of matching with an ordinary entrepreneur $\frac{1-\beta^e}{1-2\gamma\beta^e}$ multiplied by the fraction of the project's output received by the investor $(1-\delta)F_{LH}$:

$$d^i = \frac{1-2\gamma\beta^e}{1-\gamma\beta^e} \frac{1-\delta}{1-2\gamma\beta^e} [(1-2\gamma)\beta^e F_{HH} + (1-\beta^e)F_{LH}]. \quad (1.33)$$

A similar expression can be obtained for the talented entrepreneur. The probability of matching with somebody for a talented entrepreneur is equal to 1, so

$$d^e = \frac{\delta}{1-\gamma\beta^e} [(\beta^i - \gamma\beta^e)F_{HH} + (1-\beta^i)F_{HL}].$$

Due to the supermodularity of the output function, sorting is possible. There exists a separating equilibrium such that the incentive compatibility constraint for the capital-scarce investor holds (the low type has no incentive to mimic the high type). The capital-abundant investor is indifferent between being randomly matched and being matched by a banker, while the capital-scarce investor is strictly better off under random matching. In this case, the system (1.30) takes the form below:

$$\begin{aligned} p^i &= (1-\delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e &= \delta(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ (p^i + p^e) &= \delta(F_{HH} - d^i(\gamma)) + (1-\delta)d^e(\gamma). \end{aligned} \quad (1.34)$$

Surprisingly, as long as $\gamma \leq 1/2$, the income of a banker is an increasing function of the number of bankers, while the income of an entrepreneur is a decreasing function of the number of bankers. The rise of bargaining power δ has no effect on the banker's income and a positive one on entrepreneurial income. The solution of the system (1.34) is linear in δ :

$$\tilde{\gamma} = \delta \left[\frac{1}{\beta^e} + \frac{1-\beta^e}{\beta^e} \frac{F_{HH} - F_{LH}}{F_{HH}} - \frac{2(1-\beta^i)}{\beta^e} \frac{F_{HH} - F_{HL}}{F_{HH}} \right] - \frac{1-\beta^e}{\beta^e} \frac{F_{HH} - F_{LH}}{F_{HH}}. \quad (1.35)$$

There exist two thresholds $\bar{\delta} > 0$, such that $\tilde{\gamma} = 0$, and $\tilde{\delta} > 0$, such that $\tilde{\gamma} = 1/2$:

$$\bar{\delta} = \frac{(1 - \beta^e)(F_{HH} - F_{LH})}{(1 - \beta^e)(F_{HH} - F_{LH}) + (2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL}}, \quad (1.36)$$

$$\tilde{\delta} = \frac{(1 - \beta^e/2)F_{HH} - (1 - \beta^e)F_{LH}}{(1 - \beta^e)(F_{HH} - F_{LH}) + (2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL}}. \quad (1.37)$$

Depending on parameter values, both $\bar{\delta}$ and $\tilde{\delta}$ can potentially be greater than 1. The solution $\tilde{\gamma}$ exists only for $\delta \in [\min\{\bar{\delta}, 1\}, \min\{\tilde{\delta}, 1\}]$. The solution $\tilde{\gamma}$ exists as long as $\bar{\delta} \leq 1$. Using expression (1.36), the latter can be rewritten as follows:

$$(2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL} \geq 0. \quad (1.38)$$

Second, the number of bankers is greater than or equal to the number of talented entrepreneurs $\gamma \geq 1/2$. Thus, all talented entrepreneurs are matched by bankers. The number of capital-abundant investors $\beta^e(1 - \gamma)$ are matched by bankers. In this case, the solution of the Nash bargaining problem for random matching is given by

$$d^i = \frac{(1 - \beta^e)}{1 - \beta^e(1 - \gamma)}(1 - \delta)F_{LH}, \quad (1.39)$$

$$d^e = \frac{\delta}{1 - \beta^e + \gamma\beta^e} [(\beta^i - \beta^e + \gamma\beta^e)F_{HH} + (1 - \beta^i)F_{HL}]. \quad (1.40)$$

If the number of bankers is greater than the number of talented entrepreneurs, competition among bankers drives the price p^i down to zero. In this case, system (1.30) takes the form below:

$$\begin{aligned} p^i &= (1 - \delta)(F_{HH} - d^i(\gamma) - d^e(\gamma)), \\ p^e &= 0, \\ \frac{1-\gamma}{\gamma}p^i &= \delta(F_{HH} - d^i(\gamma)) + (1 - \delta)d^e(\gamma). \end{aligned} \quad (1.41)$$

As we can see, the banker's income is a decreasing function of the bargaining power δ , while entrepreneurial income is an increasing function of δ . Furthermore, the expected income of a banker grows with γ . System (1.41) can be expressed in the form of a quadratic equation in γ :

$$\begin{aligned} &\gamma^2 + \left[(1 - \delta) \frac{(1 - \beta^e)}{\beta^e} \frac{F_{HH} - F_{LH}}{F_{HH}} + \frac{\delta - \beta^e}{\beta^e} + \delta(1 - \delta) \right] \gamma - \\ & - (1 - \delta)^2 \frac{(1 - \beta^e)}{\beta^e} \frac{F_{HH} - F_{LH}}{F_{HH}} - (1 - \beta^i)(1 - \delta) \frac{\delta}{\beta^e} \frac{F_{HH} - F_{HL}}{F_{HH}} = 0. \end{aligned} \quad (1.42)$$

The solution of quadratic equation (1.42) contains two roots, but one root is always negative. For the solution to exist, the second root has to be greater than $1/2$. Let $\hat{\gamma}$ be the positive solution of (1.42). This solution exists as long as

$$\frac{\delta(1 + \beta^e(1 - \delta))}{4(1 - \delta)} \leq (1 - \beta^i)\delta \frac{F_{HH} - F_{HL}}{F_{HH}} + (1 - \beta^e) \frac{F_{HH} - F_{LH}}{F_{HH}}(1/2 - \delta). \quad (1.43)$$

Analyzing condition (1.43), we can conclude that the condition is likely to be satisfied when: the dispersion of wealth k^H/k^L is high; the share of capital-abundant investors β^i is low; and the bargaining power of entrepreneurs δ is relatively low. Furthermore, if $\delta \leq 1/2$, condition (1.43) is likely to be satisfied when the dispersion of talent is high and the share of talented agents β^e is low. When condition (1.43) is satisfied with equality, it can be rewritten as the definition of $\hat{\delta}$ defined in Proposition 2.

The constrained efficient allocation is $\gamma^* = 1/2$: This implies that ΔY given by expression (1.3) is positive and can be rewritten in the following form:

$$(2\beta^i - 1)(F_{HH} - F_{HL}) + F_{HL} < -\frac{2(1 - \beta^i)(1 - \beta^e)}{2 - \beta^e}(F_{LH} - F_{LL}). \quad (1.44)$$

The right-hand side of inequality (1.44) is negative, therefore its left-hand side is negative as well. If we compare the left-hand side of (1.44) with the left-hand side side of expression (1.38), they are exactly the same. Hence, if $\gamma^* = 1/2$, the solution $\tilde{\gamma}$ does not exist. \square

1.6.3 The solution of the decentralized equilibrium if $\gamma^* = 0$

In this section, I show the solution of the decentralized equilibrium in the case in which the constrained efficient allocation is 0. The proposition below summarizes the case:

Proposition 3. *If the constrained efficient allocation is $\gamma^* = 0$, then both equilibria with few $\tilde{\gamma}$ and many $\hat{\gamma}$ bankers are possible. In this case, there is a range of $\delta \in [\tilde{\delta}, 1]$, such that the decentralized equilibrium is constrained efficient.*

Proof: As shown in appendix 1.6.2, the solution $\hat{\gamma}$ exists as long as condition (1.43) holds, and it does not depend on whether the constrained efficient allocation

is 0 or $1/2$; the solution $\tilde{\gamma}$ does not exist if $\gamma^* = 1/2$. The solution $\tilde{\gamma}$ exists as long as $\bar{\delta} < 1$, defined by expression (1.36). The latter is likely to be satisfied if $\gamma^* = 0$. \square

1.7 Appendix B

1.7.1 Private banking and private equity finance

This subsection introduces an alternative way of calibrating the model. Most global banks, such as Credit Suisse, Barclays, BNP Paribas, Citibank, Deutsche Bank, HSBC, JPMorgan Chase and UBS, have a separate business unit with dedicated teams of client advisors and product specialists exclusively for high-net-worth individuals. They provide a wide range of investment opportunities, including bonds, stocks and more importantly private equity finance.

Private equity is an important channel through which long-term investments are made. It has grown steadily over the past three decades, and today private equity funds worldwide manage over \$1 trillion. For some countries, such as Israel, the US and the UK, private equity accounts for more than 5% of total investment (see Table 1.4 for details).

Table 1.4: The size of private equity

	% GDP		% Investment	
	2010	2011	2010	2011
Israel	0.63	2.09	3.50	10.45
UK	1.13	0.75	7.53	5.00
US	0.9	0.98	5.00	5.44
China	0.16		0.33	
World		0.30		1.58

I would like to convince the reader that the matching assumption holds for private equity. A small private equity fund provides an opportunity to invest in a

few companies over a long-term horizon for a small number of wealthy investors. As we can see from Table 1.5, private equity funds typically employ 12 professionals. These professionals select one or two companies each for the fund to invest in. Investments are large (over \$50 million). Investors are wealthy and expected to invest over a long-term horizon. The minimum required commitment rises from a median of \$1 million for funds of \$100 million or less, up to a median of \$10 million for funds of \$1 billion or more. There is no active market for private equity positions, making these investments illiquid and difficult to value. Private equity funds typically have horizons of 10–13 years, during which the invested capital cannot be redeemed. Furthermore, based on the sample of firms from the Pricewaterhouse National Venture Capital Survey, Gordon (2000) shows that 71.32% of firms reported they had received more than one offer to invest from venture capitalists. The mean number of offers was 3.18. This means entrepreneurs have a choice about which venture capitalists invest in their companies, but this choice is rather limited and matching plays an important role.

Table 1.5: Private equity funds

	N	Number of			Size (\$ mn)	
		professionals	investments	I/P	Fund	Investment
VC	94	9	20	2	225	11.25
Buyout	144	13	12	1	600	50

Given the long-term horizon and the high entry costs, the question is why investors are willing to engage in these investments. Investors are compensated well by substantially higher returns. Table 1.6 shows that the return from an investment in private equity funds is three times higher than in stocks. We can see the comparison with inflation and the returns on other assets: stocks, gold, T-bills etc.

The first important ingredient of the model is the wealth distribution. If we look at Table 1.7, we can conclude that at most 5% of households in the US can afford to invest at least \$1 million for the horizon of ten years. I target 5% as the share of capital-abundant investors.

Table 1.6: US real asset returns

Period	PEF	S&P	TBond	Gold	Inflation
1997-2011	9.2%	3.2%	0.4%	7.0%	2.4%
1975-2011		7.5%	1.3%	4.0%	4.2%

Table 1.7: Wealth distribution in the US

	1962	1983	1989	1998	2001	2007	2010
Average	194	284	326	362	468	564	464
Median	52	73	78	81	91	108	57
Top 20%	786	1157	1361	1507	1976	2397	2062
Top 5%	2120	3190	3841	4273	5542	6973	5842
Top 1%	6491	9599	12176	13650	15627	19486	16439
1% to median	125	131	156	168	173	181	288

The second important ingredient of the model is the distribution of talent. The most challenging exercise is to calibrate it. I use three different proxies. First, based on US firm-level data I estimate the moments of the productivity distribution in the US. Second, based on data on different test scores from the National Longitudinal Survey of Youth 1997 (NLSY97), I estimate the moments of the IQ distribution in the US. Finally, since my model contains only two levels of talent, we might think of these as being college graduates and high-school graduates. Therefore, I use data on the share of college graduates over time as a proxy for the increase in the supply of talent (skills).

Chapter 2

Structural Changes and Labor Income Distribution: The Importance of Educational Policies

2.1 Introduction

The relationship between inequality and growth is of fundamental interest for economists. The most challenging trends to explain of the last few decades, which have attracted the attention of both economists and policymakers, have been the growth of the service sector and the rapid increase of income inequality in developed countries. While both trends have been studied separately and extensively, economists, perhaps surprisingly, have not formally analyzed the possible interaction between the two. This paper is an attempt to do just that by shedding light on the possible interaction between income inequality and the services economy through endogenous education choice. The growth and structural change literature focuses mostly on labor demand, while the inequality and education literature emphasizes the importance of labor supply. My contribution is to simultaneously explain both trends by examining both sides of the labor market with an emphasis

on endogenous education choice.

Before going into detail, I consider income inequality and services separately. The growth of services and the decline of manufacturing are well known and well documented. However, the service sector is large, up to 70% of GDP, and very heterogeneous. It can be roughly split into two subsectors, which are very distinct in multiple dimensions: producer services and consumer services. Consumer services include personal services, transportation, entertainment, retail and wholesale trade. Producer services include business services, telecommunications, finance, insurance and real estate. While consumer services are mostly used for final consumption and demand low-skilled labor, producer services are mostly used as intermediate goods and demand high-skilled labor. Manufacturing is very similar to consumer services in terms of skills and the stage of consumption. Furthermore, the employment share of consumer services has remained roughly stable over the 20th century in the US, while the employment share of producer services has almost quintupled at the expense of manufacturing.

Many authors have observed the growth of labor income inequality in developed countries, in particular, the increase of wage inequality between educational groups. Despite an increasing supply of college graduates, the college wage premium, the additional average salary a college graduate earns relative to a high-school graduate, has increased sharply in the US from 40% in the 1950s to 90% in the late 2000s (see Figure 2.4). The study by Autor et al. (2008) shows that the effect of the increased college wage premium can explain up to 50% of the rise in overall labor income inequality. To summarize, empirical evidence suggests that the accumulation of skills plays an important role in explaining both trends: the growth of services, mostly producer services, and the increase of income inequality.

Employing skill-biased technical change in a multi-sector framework as a standard mechanism to generate the growth of the college wage premium, this paper concentrates on labor supply by analyzing an endogenous benefit, the college wage premium, and an endogenous cost, a tuition fee. Tuition fees in the US have grown faster than median incomes and even the college wage premium over the last few decades, which might limit access to university and consequently high-skilled la-

bor supply. (See Figure 2.8 for the US. Greenaway and Haynes (2003) document a similar trend for the UK.)

This paper studies how an endogenous education choice can affect the economy's sectoral composition and income inequality. It analyses two possible ways of financing education: a bequest and a loan. In the first way, which might be seen as traditional, parents leave educational bequests that determine the occupational choices of their children. Occupational returns are determined by market conditions. In the second way, which can be seen as modern, individuals employ loans as the main source of funding for university tuition fees.

I develop a model with heterogeneous agents who differ in ability, who face an educational cost and have to decide whether to study or not. The model is a three-sector general equilibrium model, in order to better understand the driving forces and interconnections between the rise of the service sector and wage inequality. The first sector is producer services, which only demands high-skill labor and is used as an intermediate input for manufacturing; the other two are consumer services and manufacturing, which both compete for unskilled labor. Both sectors provide goods for final consumption. The share of services and income inequality are functions of the wage premium and the ratio between skilled and unskilled labor, which are determined endogenously in the model.

As mentioned, the different ways of financing education have different implications for efficiency, as well as for sectoral growth and inequality. The model with inheritance predicts an endogenous and permanent separation of the population between the rich (skilled) and the poor (unskilled). Due to the absence of financial constraints, the model with loans leads to an efficient outcome, even though it may still generate a persistent level of inequality.

The model and data show that the growth of services and the rise of wage inequality can be linked through the fraction of college graduates and the college wage premium. As we will see below, the data suggests that the growth of services in the US is mostly accounted for by producer services. These services demand mainly high-skilled labor. As producer services grow, due to higher productivity, for example, they demand more and more high-skilled labor. To attract labor, pro-

ducer services increase wages. The increase of wage inequality between educational groups has a very important impact on the growth of overall income inequality. As a result, we observe both the rise of services and the increase of income inequality. However, as pointed out above, higher inequality may in some cases limit access to college for middle-class and poor households, and consequently restrict the supply of high-skilled labor.

The paper is structured as follows. The next section reviews related literature. Section 3 then provides empirical motivation for the current research and data. The model and theoretical results are described in Section 4. The last section concludes and motivates further research.

2.2 Related Literature

My paper is related to a huge range of literature, of which I will provide a very limited review. The review can be split into human capital growth literature, education choice literature, and literature related to structural changes and R&D.

Let me first review the branch of the literature that deals with structural changes and R&D. The rise of services has at least three theoretical explanations: a shift of demand towards services; international trade and outsourcing; and inter-sector productivity differences. The first explanation recalls the so-called “hierarchy of needs” hypothesis proposed by Clark (1940). The hypothesis states that as income grows, a higher share of household expenditure will be spent on services. This result can be easily reproduced by using non-homothetic preferences (Echevarria, 1997; Buera and Kaboski, 2011). The second explanation relies on international trade, which may lead to specialization in skill-intense industry in developed countries and could affect the demand for skilled labor.¹ The last explanation, based on the fact that different sectors experience different productivity growth, is used in this paper.² Surprisingly, Buera and Kaboski (2009a) demonstrate that neither non-homothetic preferences nor skill-biased technical change

¹See Acemoglu(2003), Blum(2008).

²Baumol (1967, 2001) for empirical evidence of the rise of services and Ngai and Pissarides (2007) for a theoretical explanation.

can quantitatively explain the recent structural changes in the United States.

As discussed in the introduction, there is consensus in the literature that technological change has been skill-biased for several decades and has generated an increasing wage premium. However, this result is no longer true in the multiple sectors model, because wages depend not only on productivity levels, but also on prices of goods (Weiss, 2008). A similar observation about the importance of the multiple sectors model was made originally by Baumol (1967).

The second branch of literature deals with an educational choice with or without credit market imperfections, and how this choice affects the income distribution. Some papers argue that credit market imperfections do not limit access to education,³ while others claim the opposite (Kane, 1994). De Fraja (2002) showed that in the presence of a credit constraint government can restore an efficient outcome by public intervention in the provision of education. Moreover, Becker (1964) advocates government subsidies of educational loans, because of the difficulty of using human capital as collateral. More importantly, even in the absence of credit constraints, a market might produce persistent inequality.

Many economists who conduct research in the field of education have contrasted public and private education systems and studied their inequality implications. The standard result is that the public education system leads to a reduction of inequality, while private education might have a positive impact on income inequality.⁴

To the best of my knowledge, there is no paper that formally links income inequality and a service sector. However, I would like to refer to two of the most relevant papers, which are Eicher (2001), who considered a static model of the labor market with non-monotonic demand for skilled labor, and Buera and Kaboski (2011), who examined the importance of high-skilled labor in the rise of the service sector.

³See Cameron and Heckman (2001), who provided a detailed analysis of educational attainment and college participation in the US and concluded that the importance of short-term credit constraints is greatly exaggerated.

⁴In Cardak (2004), parents can choose between public or private schools and can vote on taxes used to fund public schools. See also Glomm and Ravikumar (1992).

Buera and Kabosky's paper presents the consumption demand-driven story, which relies on three elements: an exogenous distribution of productivity growth, a special type of non-homothetic preferences and a convex schooling cost. As productivity increases, it generates a rise in the college wage premium, which is completely exogenous to the labor supply decision. The increasing premium affects income and, consequently, stimulates demand through non-homothetic preferences. The production of market services requires high-skilled labor and intermediate goods, while the production of home services demands low-skilled labor and intermediate goods. As productivity rises exogenously, an individual's consumption moves into even more complex wants, eventually those in which high-skilled labor has a comparative advantage. The representative household decides what fraction of the household should be specialized high-skilled workers and what fraction should be low-skilled workers, and ensures an equal level of consumption for all members of the household.

Even though the set-up of my paper seems similar to Buera and Kabosky's paper, there are significant differences. I would like to emphasize several main differences: an endogenous educational choice, a basket of final consumption items and inequality implications. The first difference is that an endogenous educational choice induces interaction between labor supply and labor demand. The second difference is that, in Buera and Kabosky's paper, final consumption takes the form of services, while in my model, a basket of goods and services represents final consumption. The last difference is that, since there is a representative household, Buera and Kaboski (2011) do not explicitly discuss an application of their model to income inequality, but it is possible to understand the behavior of income inequality from their results. The college wage premium explodes, while the relative share of high-skilled workers converges to a constant level. As a result, income inequality also explodes.

Eicher and Garcia-Penalosa's paper provided an alternative mechanism based on the dual role of human capital to generate simple intra-sectoral inequality. However, I believe that Eicher and Garcia-Penalosa's results can be interpreted in a different way. Because the service sector hires mainly skilled labor, while

manufacturing demands unskilled labor, as a country accumulates skills (human capital), we observe a shift of employment towards services. As a result, it is essentially the same intra-sectoral inequality mechanism. However, the authors did not discuss the application of their model to sector decomposition, even though it was feasible in their framework, because the model operates in terms of skilled labor and unskilled labor and the wage premium. The second main difference is that I shift the focus onto the labor supply side of the story.

2.3 Data

This section justifies the assumptions of the model and formally links services to the share of skilled workers and income inequality to the wage premium.

My paper attempts to show that there is a nontrivial relationship between income inequality and sectoral composition. In particular, the post-industrial society may be characterized by not only a relatively large service sector, but also by relatively high income inequality. If we look at US data (Figure 2.1), we observe that US society was as unequal at the beginning of the century as it is now. Income inequality dropped dramatically during the 1940s and reached a minimum in the 1950s; then it started to grow. The ratio of value added in services to manufacturing showed the same pattern. The minimums of the two graphs are quite close to each other (less than 10 years difference or 14% of GDP per capita). These U-shaped relationships can be viewed through the migration of the workforce from manufacturing, with low value-added, to high value-added ‘producer’ services.⁵ However, there is an important difference, which is that low-skilled labor from industry is not a perfect substitute for high-skilled labor, which is required to produce producer services.⁶

⁵See Figure 2.6.

⁶This dynamic is not specific to the US. Figures 2.2 and 2.3 present panel data of the value added in services to manufacturing against the natural logarithm of real GDP per capita in \$ (in 1990 prices) and the top 5% income share against GDP per capita (in 1990 prices), respectively. The data demonstrates a clear U-shaped pattern with a minimum around per-capita income of \$7,100 and definite growth starting after \$12,000. A similar U-shaped relationship can be

As we have seen, the growth of services started before the rise of income inequality. This time lag can be explained by two facts. First, in the seventies and eighties the US economy experienced structural changes, which were associated with a decline in the wage premium (Acemoglu, 2002). Second, Card and Lemieux (2001) discovered that the entire rise of the wage premium can be attributed to changes in the earnings of relatively young college-educated workers. The college–high school wage gap for younger men with 3–7 years of experience doubled over the 30 years between 1959 and 1989, whereas the gap for older men remained nearly constant. Furthermore, the profile of the wage premium for men peaks at about 13 years of experience. As a result, as services grow, the demand for educated workers increases, but the rise of earnings happens with a time lag, which we observe in Figure 2.1.

While the college wage premium has shown a U-shaped trajectory over time, the relative supply of skills has constantly increased. Figure 2.4 shows the US experience over the past century.⁷ The growth has accelerated since the 1970s and has been accompanied by a sharp increase in the college premium.⁸ Based on widely available data for the US, I argue that Figures 2.1 and 2.4 present fundamentally the same phenomena, i.e. the rise of services and the increase of income inequality.

Having discussed the benefits of being college educated, let me turn to a cost. Figure 2.5 shows that the cost of higher education, which is mostly tuition fees, remained stable till the 1950s and then started to grow faster than median income and even the college wage premium. Tuition fees at Harvard university are used as a proxy for average tuition fees because, unfortunately, the long-term data for

observed in Figure 2.3, where the top 5% income share is plotted against GDP per capita, but the minimum of the graph appears to be around \$15,000 and definite growth starts after \$20,000.

⁷Since the data of college wage premium is combined from different sources (Goldin and Katz, 2007; Acemoglu and Autor, 2011), caution should be taken.

⁸A similar, but less marked trend has been discovered for other countries with some exceptions. When Walker and Zhu (2008) estimated the college wage premium for the UK between 1994 and 2006, they found that although the dramatic increase in higher education enrolment, the college wage premium remained stable for men and rose for women.

average tuition fees are not available.⁹ As we see from Figure 2.5, tuition fees at Harvard are a good proxy for the average tuition fees in a private university. Tuition fees in the US have been growing dramatically and soon even a public college will not be affordable for 20% of Americans, while a private college will become unaffordable for half of Americans. The New York Times commented on the current situation: “The rising cost of college – even before the recession – threatens to put higher education out of reach for most Americans. . . Over all. . . student borrowing has more than doubled in the last decade” (Lewin, 2008). Moreover, Figure 2.9 shows that the average loan per full-time equivalent has risen dramatically over a few decades. In 2008, total college debt exceeded total credit card debt.

Moving from the individual part of the story to services data, services are the biggest sector in every developed country and they are very diverse. To examine them, we need to decompose them into smaller parts. According to the standard broad industry group classification, services include the following industries: Utilities, Transportation, Telecommunications, Wholesale Trade, Retail Trade, Finance, Insurance, and Real Estate (FIRE), Business and Repair Services, Personal Services, Entertainment and Recreational Services, Social Services, and Government. To connect services to high-skill labor, let me point to Table 2.1, which presents the fractions of college-educated labor, total man-hours and value added in different industries in 1940 and 2000. The data clearly shows that there is a significant difference in terms of the fraction of college graduates in the labor force between services and manufacturing. In 2000, the fraction of college graduates in services was around 60%, while less than 40% of workers in manufacturing had a college degree.

Following Buera and Kaboski (2011), Table 2.1 presents the ranking of services into two groups: low-skill and high-skill. High-skill industries are those industries with at least 60% of college-educated labor in 2000. High skill services include: Telecommunications, FIRE, Business and Repair Services, Social Services, and Government, while low-skill services with minor exceptions embrace: Trans-

⁹The tuition fees data for the University of Pennsylvania and the University of Iowa demonstrate the same pattern.

portation, Wholesale Trade, Retail Trade, Personal Services, Entertainment and Recreational Services.

Broadly speaking, the high-skill services can be subdivided into two categories: services that are related to production goods ('producer' services), such as telecommunications, FIRE, business and repair services, and services that are strongly influenced by government policies, such as the majority of social services and public administration. My paper focuses on production-related services, which, in 2000, accounted for 16% of total man-hours and 36% of value added (See Table 2.1). All 'producer' services have high information content or even require a specialized degree, for example law, accountancy or risk management. Apart from the government sector, the two main industries of social services, which are healthcare and education (18% of the total labor force), whose growth may be driven at least in part by growth in government subsidies or other policies. To avoid possible effects of government policies, the government-related services are excluded from the current analysis.

Comparing the industries in Table 2.1, several important aspects should be mentioned. First, the ranking of industries is remarkably stable, i.e. the higher the fraction of college-educated people was in 1940, and the higher it is now. Second, the high skill-intense industries grew not only extensively in terms of persons engaged in production by industry, but also intensively in terms of value added per employee (see Figures 2.5 and 2.6). The values in Table 2.1 and Figure 2.6 are different, because they are based on different sources of data and operate in different terms. Third, although the fact that low-skill services had approximately the same fraction of college graduates as non-services producer over time, the share of low-skill services production only slightly decreased in comparison with manufacturing and agriculture, which dropped dramatically. Furthermore, Figures 2.5 and 2.6 clearly show that the growth of services is mostly driven by producer services and slightly by government and social services. Another important feature of the data is that the non-service sectors shrank dramatically, while the low-skill services declined slightly, but steadily. Last, both the decrease of the non-services share and the increase of the 'producer' services share decelerate. Perhaps the share

of GDP components may stabilize around current levels. Otherwise, it would be strange to observe an unlimited growth of ‘producer’ services, while the industrial sector declines. However, not only the real sector demands ‘producer’ services.

The last important feature of the US services economy is that so-called ‘producer’ services provide output for intermediate consumption as well as for final consumption. Therefore it is not fully correct to call FIRE, legal, communication and other business services the ‘producer’ services, because, by definition, ‘producer’ services are intermediate inputs to further production. It is conventional wisdom to consider FIRE, legal, communication and other business services as ‘producer’ services, but this is only partly true. The truth is that all so-called ‘producer’ services provide both intermediate and final household consumption, and the share of intermediate products varies greatly (see Table 2.2). For example, in the US, the FIRE sector in fact provides more than 50% for final consumption, while almost 90% of the R&D sector is intermediate input. However, on average, in 2005, producer services provided 40% of total output for final consumption, which is higher than the average for the whole economy, and only 53% for intermediate inputs. Moreover, the data suggests that the distribution between final and intermediate consumption is roughly stable over time.

To summarize, the data presented in this section clearly shows that the rise of services has been driven by so-called ‘producer’ services,¹⁰ which hire mainly high-skill graduates and pay them substantially higher salaries. The producer services have grown not only intensively, but also extensively, but at present they are still less than 20% of the labor force. To attract more labor, the FIRE sector (the main part of ‘producer’ services) increased wages. The growth of wages has had a direct impact on labor income inequality, which is now approximately equivalent to income inequality. The rise of the wage premium has been accompanied by the even faster growth of the cost of education; therefore the logical implication is that the cost of education will limit the rise of college enrolment.

As a result, this data section provides strong evidence to support two claims: the rise of the service sector is directly related to the increase of the fraction of

¹⁰40% of ‘producer’ services output accounts for final consumption.

college graduates, and the rise of the wage premium has a direct impact on income inequality. The next section develops the model that links income inequality and the service sector.

2.4 Model

As mentioned, the model includes the following components: multiple sectors, growth, an endogenous education choice and distribution of abilities. To study structural changes, it is necessary to have multiple sectors. Apart from this fact, a multiple sectors model allows us to explore important general equilibrium effects. In particular, relative wages depend not only on productivity levels, but also on prices of goods. This can reverse results (Weiss, 2008). Both exogenous and semi-endogenous productivity growth are analyzed and contrasted in the paper. The main component is an endogenous educational choice.

There are two different models in my paper, because I consider two ways of financing higher education: through loans and bequests. In the first case, after observing their ability and tuition fees, individuals decide to borrow or not to finance higher education. In the second case, parents leave an educational bequest to their children without knowing their ability. The bequest as well as the loan can be spent only on education. Different ways of financing higher education affect only the household side of the story (labor supply), therefore the production side remains the same for both models. The next subsection provides a description of production.

The model is intended to be as simple as possible to have a closed-form solution, while allowing for the potentially complicated effects outlined in the empirical part. The analysis focuses on steady states and transitional dynamics. Substantial emphasis is put on the changes in labor supply due to different education policies. Data suggests that the service sector hires a significantly higher share of skilled labor in comparison with manufacturing. As the service sector increases, demand for skills increases as well. It forces the skilled wage premium to rise as well, but the cost of education grows differently for different individuals.

2.4.1 Production

My model includes three sectors, which are producer services, low-skill services and manufacturing. The producer services sector hires high-skilled labor (college graduates) and produces intermediate goods. The output of this sector together with low-skilled labor is used in the manufacturing sector to produce one of the final consumption goods. The last sector is a low-skill service that uses only low-skilled labor to produce the second final consumption good. All three sectors are competitive, therefore the zero-profit condition holds.

The production of producer services requires only one input, which is low-skilled labor, and demonstrates decreasing returns to scale. The high-skill services firm's maximization problem is given by

$$\max_{L_s} [p_s q_s L_s^\alpha - w_s L_s].$$

FOC:

$$w_s = \alpha p_s q_s L_s^{\alpha-1}. \quad (2.1)$$

Equation (2.1) represents the first-order condition, where $\alpha \in (0, 1)$ is a parameter; w_s and q_s are the wage and the productivity level in producer services; and p_s is the price of producer services. As discussed, the wage depends not only on the productivity level, but also on the price of the good. L_s is total demand for high-skilled labor.

The manufacturing production technology is described by the CES function, which admits some degree of substitutability $\sigma \in (0, \infty)$ between intermediate inputs and high-skilled labor. The manufacturing firm's maximization problem is given by

$$\max_{L_m, Q_s} [p_m ((q_m L_m)^{(\sigma-1)/\sigma} + Q_s^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)} - w_m L_m - p_s Q_s].$$

FOC:

$$w_m = p_m q_m^{(\sigma-1)/\sigma} L_m^{-1/\sigma} (q_m L_m)^{(\sigma-1)/\sigma} + Q_s^{(\sigma-1)/\sigma} 1/(\sigma-1). \quad (2.2)$$

$$p_s = p_m Q_s^{-1/\sigma} (q_m L_m)^{(\sigma-1)/\sigma} + Q_s^{(\sigma-1)/\sigma} 1/(\sigma-1), \quad (2.3)$$

where $Qs = q_s L_s^\alpha$ is the output of producer services; w_m and q_m are the wage and the productivity level in manufacturing; and p_m is the price of manufacturing.

The maximization problem of the consumer (low-skill) services firm is expressed by

$$\max_{L_l} [p_l q_l L_l^\alpha - w_l L_l].$$

FOC:

$$w_l = \alpha q_l p_l L_l^{\alpha-1}. \quad (2.4)$$

Equation (2.4) represents the first-order condition, where $\alpha \in (0, 1)$ is a parameter; w_l and q_l are the wage and the productivity level in low-skill services; and p_l is the price of low-skill services. Competition between manufacturing and low-skill services for low-skilled labor leads to equal wages in both sectors: $w_l = w_m$.

Equations (2.1)–(2.3) taken together lead to the equation for labor demand:

$$\omega = \alpha (q_s/q_m)^{(\sigma-1)/\sigma} L_s^{\alpha-1-1/\sigma} L_m^{1/\sigma}, \quad (2.5)$$

where $\omega = w_s/w_m$ is the college wage premium depending on the relative productivity level and the relative demand for skilled labor.

Expressions (2.3) and (2.4) taken together lead to the expression for the relative price of final consumption goods:

$$p = \alpha q_l L_l^{\alpha-1} (q_m)^{(1-\sigma)/\sigma} L_m^{1/\sigma} [(q_m L_m)^{(\sigma-1)/\sigma} + Q_s^{(\sigma-1)/\sigma}]^{1/(1-\sigma)}. \quad (2.6)$$

As discussed, I consider two types of productivity process: exogenous and semi-endogenous growth. I follow the seminal paper by Jones (1995):

$$(q_s)_t = (1 + g)(q_s)_{t-1}, \quad (2.7)$$

$$(q_s)_t = (q_s)_{t-1} + (q_s)_{t-1}^\mu (L_s)_{t-1}^\gamma. \quad (2.8)$$

The last condition, which needs to be satisfied, is sectoral allocation given by:

$$L_s + L_m + L_l = 1. \quad (2.9)$$

2.4.2 Model with Bequests

I employ the standard overlapping generations (OLG) framework, adopted by Glomm and Ravikumar (1992) and developed by Galor and Moav (2004). There are two types of agents: parents and their children. The agents live two periods and they are differentiated by abilities and initial asset holdings. There is an exogenous distribution of the people in terms of ability that is characterized by a density function $\psi(a)$.

In every period the economy is populated by children and their parents. In the first period, the children only study to be high-skill employees during the next period if their parents decide to pay tuition fees, but not otherwise. In the second period, parents earn money depending on their educational level and divide it between their own consumption and tuition fees for their children. I employ this approach, because the equilibrium concept is simple and allows us to have a closed-form solution. Thus individuals simply allocate their wealth optimally between their own consumption and their bequests to their offspring (tuition fees for their children).¹¹

Hence the income of the agents is predetermined by the decision of the parents in the previous period. There is a positive wage gap between college-educated employees, who work in the service sector, and school-graduated workers, who are employed in the production sector. However, there is a cost of being a college graduate, which is the tuition fee. The cost of higher education is an increasing

¹¹There are at least four formulations of the intergenerational link in the distribution and growth literature. Three of them deal with altruism in the sense that some bequest to children is directly incorporated into the utility function of the parents, in contrast to a life-cycle model with inheritance. First, members of the current generation could value the utility level achieved by their descendants (Loury, 1981). Second, they could value the allocations of their descendants (Kohlberg, 1976). Third, members of the current generation may value the wealth they pass on to their descendants (Banerjee and Newman, 1991). Finally, there is the Blanchard–Yaari approach of perpetual youth (Blanchard, 1985; Yaari, 1965), which is a standard OLG model with the key additional assumption that an agent has a constant probability of death, independent of age. The dead agent is replaced by a newborn agent who inherits his wealth as well (Blanchard and Fischer, 1989). In the first two formulations, members of each generation take as given the optimal decision rules of their offspring.

function of median income and a decreasing function of ability and the number of skilled people, which plays a critical role in the model, because it prevents all people from being college educated.

In period t , given earnings E_t , which are predetermined by the parents' decisions b_{t-1} and the level of ability a_{t-1} , the adult household maximizes its own levels of consumption $(c_m)_t$ and $(c_l)_t$ and the amount of bequest b_t to its children. The maximization problem of a household is:

$$\begin{aligned} & \max_{c_m, c_l, b} [\log(\mathbf{c}_t) + \theta \log(b_t)] \\ & \mathbf{c}_t + b_t \leq E_t(b_{t-1}, a_{t-1}) \\ & \mathbf{c} = (\eta_m(c_m)^{(\epsilon-1)/\epsilon} + \eta_l(c_l)^{(\epsilon-1)/\epsilon})^{\epsilon/(\epsilon-1)} \\ & a \in \Phi(.) \end{aligned}$$

where $\theta \in [0, 1]$ is a measure of parental altruism. Due to the log-log utility specification, all variables are defined in relative terms with respect to the wage of a low-skill individual $(w_m)_t$. Then, the earnings can be either ω if the adult individual is a college graduate, or 1 otherwise.

$$E_t = \begin{cases} \omega_t & \text{if } b_{t-1} > \tau(a_{t-1}, \omega_{t-1}) = \bar{\tau} \frac{\omega^{1/\rho_1}}{a^{1/\rho_2}} \\ 1 & \text{otherwise} \end{cases} \quad (2.10)$$

Solving for the first-order conditions of the household problem, we find the following:

$$\begin{aligned} b_t &= \begin{cases} \frac{\omega_t \theta}{\theta+1} & \text{if } b_{t-1} \geq \tau(a_{t-1}, \omega_{t-1}) \\ \frac{\theta}{\theta+1} & \text{otherwise} \end{cases} \\ \mathbf{c}_t &= \begin{cases} \frac{\omega_t}{\theta+1} & \text{if } b_{t-1} \geq \tau(a_{t-1}, \omega_{t-1}) \\ \frac{1}{\theta+1} & \text{otherwise} \end{cases} \\ \frac{c_m}{c_l} &= \left(\frac{\eta_m p_l}{\eta_l p_m} \right)^\epsilon \end{aligned}$$

Due to homotheticity of preferences, it is easy to derive aggregate earnings \mathbf{E} and consequentially, using expression (2.10) the aggregate spending on education pB_m and the total consumption expenditure $\mathbf{C} = C_l + pC_m$. These are given by

$$\mathbf{E} = w_l(L_s\omega + (1 - L_s))$$

$$\begin{aligned} pB_m &= \frac{\theta}{(\theta+1)} \frac{1}{p_l} \mathbf{E} \\ C_l + pC_m &= \frac{1}{(\theta+1)} \frac{1}{p_l} \mathbf{E} \end{aligned}$$

The next period labor supply is determined by the current period decision and given by

$$(L_s)_t = (L_s)_{t-1}(1 - \Phi(\bar{a}^1(\omega_{t-1})) + (1 - (L_s)_{t-1})(1 - \Phi(\bar{a}^2(\omega_{t-1}))) \quad (2.11)$$

There are two thresholds in terms of ability, given by:

$$\begin{aligned} \bar{a}^1(\omega) &= \frac{\omega^{\rho^2/\rho^1}(\theta+1)^{\rho^2}}{\omega^{\rho^2}\theta^{\rho^2}} \\ \bar{a}^2(\omega) &= \frac{\omega^{\rho^2/\rho^1}(\theta+1)^{\rho^2}}{\theta^{\rho^2}} \end{aligned} \quad (2.12)$$

Equation (2.11) represents the important dynamics of labor supply. The whole population is split into two groups: the high-skilled workers (the rich) and the low-skilled workers (the poor). The first element describes the share of people from rich families, who acquire higher education, while the second element represents the share of people from poor families. As we see from equations (2.12), for a given college wage premium ω , the threshold for rich people \bar{a}_{t-1}^1 is significantly lower than the threshold for poor people \bar{a}_{t-1}^2 . Contrary to the standard result of a bequest model, which is the division of population into stable dynasties of the rich and the poor, my model generates mobility between educational groups. To put it differently, a very talented student from a poor family can acquire higher education and become rich.

It is clear that the economy with educational bequests leads to an inefficient allocation, due to the fact that the occupational choices of children are determined not only by the level of ability, but also by the size of educational bequests. Since agents cannot borrow, they cannot overcome this issue. As a result, there is room

for government policies. The two most obvious policies are redistributive taxation from the rich to the poor and educational loans. The first of these is not in the scope of the current analysis, while the second will be discussed in detail in the next subsection.

2.4.3 Model with Loans

The current section presents a model with fixed-rate government-guaranteed loans. As opposed to the previous model, now individuals can borrow as much as they want to borrow. As mentioned above, in the absence of financial constraints, loans should lead to an efficient outcome. However, there still exists a small source of inefficiency, which is the fixed interest rate. The government-guaranteed loan was chosen instead of a market loan, because in regard to higher education this choice is an accurate description of reality.¹²

The maximization problem of a household is:

$$\begin{aligned} & \max_{c_m, c_l} [\log(\mathbf{c})] \\ & \mathbf{c} = (\eta_m(c_m)^{(\epsilon-1)/\epsilon} + \eta_l(c_l)^{(\epsilon-1)/\epsilon})^{\epsilon/(\epsilon-1)} \\ & \mathbf{c}_t + (1+r)d_{t-1} \leq E(a_t)_t \\ & a \in \Phi(\cdot) \end{aligned}$$

Comparing the current maximization problem with the maximization problem from the previous subsection, we can notice that parents are not altruistic anymore towards their children, which corresponds to the case of $\theta = 0$, but now children can borrow without a constraint. The earnings of the children depend on their occupational choices, which depend on their abilities. As a result, the economy with unconstrained borrowing can reproduce an outcome close to the efficient one. Since the government supplies loans, it needs to balance its budget by lump-sum transfers or taxes tr_t . The earnings of an individual are given by

¹²In 2011, less than 5% of the loans were issued by private lenders. The estimated size of the private student loan market was \$6 billion, compared with Federal Student Aid, which delivered \$157 billion of federal aid, mostly in the form of loans. Source: the College Board.

$$E = \begin{cases} \omega_t + tr_t & \text{if } d_{t-1} > 0 \\ 1 + tr_t & \text{otherwise} \end{cases}$$

Solving for the first-order conditions of the maximization problem, we find the following:

$$d_{t-1} = \begin{cases} \tau(a, \omega_{t-1}) & \text{if } \omega_t - (1+r)\tau(a, \omega_{t-1}) \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.13)$$

$$\frac{c_m}{c_l} = \left(\frac{\eta_m p_l}{\eta_l p_m} \right)^\epsilon$$

There is a common threshold for all individuals \bar{a}_t . Compare to equations (2.12).

$$\bar{a}_t = \frac{\omega_{t-1}^{\rho^2/\rho^1} (1+r)^{\rho^2}}{(\omega_t - 1)^{\rho^2}} \quad (2.14)$$

Then aggregate labor supply, in contrast to equation (2.11), is given by:

$$(L_s)_t = (1 - \Phi(\bar{a}_t)) \quad (2.15)$$

As mentioned, the government should have a balanced budget, which means that the current generation's demand for loans should be covered by the revenues from the previous generation of borrowers. If not, the surplus/deficit of the government is redistributed/covered by taxes.

$$(1+r) \int_{\bar{a}_{t-1}}^{\infty} \tau(a, \omega_{t-1}) \phi(a) da - \int_{\bar{a}_t}^{\infty} \tau(a, \omega_t) \phi(a) da = tr_t \quad (2.16)$$

Aggregate earnings \mathbf{E} and consequentially, using (9) aggregate spending on education pB_m and total consumption expenditure $\mathbf{C} = C_l + pC_m$ are given by

$$\mathbf{E}_t = (w_l)_t ((L_s)_t \omega_t - \int_{\bar{a}_t}^{\infty} \tau(a, \omega_t) \phi(a) da + (1 - (L_s)_t)) \quad (2.17)$$

$$\begin{aligned} pB_m &= \frac{\theta}{(\theta+1)} \frac{1}{p_l} \mathbf{E} \\ C_l + pC_m &= \frac{1}{(\theta+1)} \frac{1}{p_l} \mathbf{E} \end{aligned} \quad (2.18)$$

2.4.4 Income Inequality

We now turn to the impact of an increasing share of high-skill labor over time on income inequality and the share of services. It is obvious that the employment share of producer services $L_s/(1 - L_s)$ is the relative supply of college skill, and therefore an increasing function of L_s . On the other hand, labor inequality I can be expressed as follows:

$$I = (1 - L_s)L_s(\omega - 1)^2.$$

To understand the direction of changes in income inequality over time we compute the following expression:

$$\frac{\Delta I}{\Delta t} = (\omega - 1) \left[(1 - 2L_s)(\omega - 1) \frac{\Delta L_s}{\Delta t} + 2(1 - L_s)L_s \frac{\Delta \omega}{\Delta t} \right]. \quad (2.19)$$

The first part of expression (2.19) is growing over time if L_s is growing over time up to the point at which $L_s = 0.5$. The second part of expression (2.19) increases over time if ω increases over time. As we can see later on, L_s is growing over time according to both the model and the data, but it is below 0.5. The college wage premium exhibits U-shape dynamics and generates the U-shape dynamics of income inequality.

2.5 Simulations and Results

This section provides the results for four specifications of models: the bequest economy with exogenous growth and semi-endogenous growth of high-skilled labor productivity, and the education loans economy with exogenous growth and semi-endogenous growth of high-skilled labor productivity. First, as mentioned above, the loans economy produces a more efficient allocation than the bequest economy.

Second, there is almost no difference between semi-endogenous and exogenous growth.

The first set of results, related to allocation and wages, is presented in Figure 2.10. Only one sector experiences a technological improvement, which is the producer services sector. Since the output of this sector Q_s and low-skilled labor in manufacturing L_m are gross substitutes, there is no need in L_m that makes L_m converges to zero. Following labor demand equation (2.5), the decline of L_m compensates for the growth of relative productivity and stabilizes the college wage premium ω at a constant level in both cases: exogenous and semi-endogenous growth. The constant college wage premium translates through labor supply equation (2.11) in the case of the bequest economy or equation (2.15) in the case of the loans economy to the constant level of high-skilled labor L_s . Hence, to satisfy sectoral allocation, the labor in low-skill services should converge to a constant level as well.

The second set of results, related to consumption expenditures and tuition fees, is presented in Figure 2.11. The two final consumption goods are gross complements, so the ratio between them in final expenditure should converge to a constant level due the adjustment of their relative price. The bequest economy spends relatively more on services than the loans economy due to the fact that the bequest introduces inefficiency in terms of labor allocation. To put it differently, a significantly higher share of the labor force works in low-skill services than it should. Tuition fees per student and average expenditures on education converge to constant levels as well.

Since the direct comparison of two economies can be misleading due to the fact that individuals in these economies have different preferences, I decide to put the two models in a historical perspective and to see what results can be generated by switching from the bequest economy to the loans economy. Traditionally, higher education was financed by intergenerational transfers from parents to their children, which is a bequest motive. Then, more recently, loans were introduced. As we look at the data, we observe that on all graphs the 1950s marked an important turning point. The college wage premium (Figure 2.4) and the ratio between

services and manufacturing reached their minimums; tuition fees and the relative supply of college skills kept constant till the 1950s and grew rapidly thereafter. What important event happened in the 1950s?

My answer is that US Government-backed student loans were first introduced in the 1950s under the National Defense Education Act. The student loan program was established in response to the Soviet Union's launch of the Sputnik satellite, and a widespread perception that the United States was falling behind on education and technology in the Cold War. Given this fact, the 1950s was the turning point, when the US economy was shifted from the traditional way of financing education to the modern one.

Confronting the simulated graphs (Figure 2.12) and the data (Figure 2.4), we see that a “regime-switching” story can explain not only the drop in the 1950s and the further growth of the college wage premium, but also the piecewise broken growth in the share of college graduates. The current research can also explain the experiences of major developed economies in the world, commonly referred to as the “productivity slowdown”, during the same period when the college skill premium soared.

By comparing Figure 2.13 and Figure 2.8, we can see that the historical model can also account for the piecewise broken growth of tuition fees and personal expenditure on higher education, and the rise of loans per student.

Even though the model fails to explain the full dynamics of labor shares in different sectors (compare Figure 2.14 and Figure 2.6), it generates the correct behavior of the ratio between labor shares in manufacturing and services (see Figures 2.7 and 2.15).

2.6 Conclusion

This paper develops a set of general equilibrium models of the service economy with heterogeneous agents. Placing these models into historical perspective and considering bequests as the traditional way to finance higher education, and loans as the modern one, the modeled economy qualitatively reproduces the dynamics

of skill supply, the college wage premium, tuition fees and the labor allocation between sectors over the 20th century by switching from bequests to loans.

Of course, I make no claim that the only possible mechanism of interaction between income inequality and services is through the wage premium and the share of skilled labor, but I do believe that this mechanism is important.

It has been argued that high-skill services are distinct enough from the other types of services to be considered independently. These services evolve in well-developed countries and require a highly educated workforce. Since technological changes (Acemoglu, 2002; Aghion, 2002) seem to be skill biased, this type of workforce can demand high salaries that influence wage inequality.

Bibliography

- ACEMOGLU, D. (2002): “Directed Technical Change,” *Review of Economic Studies*, 69, 781–809.
- ACEMOGLU, D. AND D. AUTOR (2011): *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Elsevier, vol. 4 of *Handbook of Labor Economics*, chap. 12, 1043–1171.
- AGHION, P. (2002): “Schumpeterian Growth Theory and the Dynamics of Income Inequality,” *Econometrica*, 70, 855–882.
- AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): “Trends in U.S. Wage Inequality: Revising the Revisionists,” *The Review of Economics and Statistics*, 90, 300–323.
- BANERJEE, A. V. AND A. F. NEWMAN (1991): “Risk-Bearing and the Theory of Income Distribution,” *Review of Economic Studies*, 58, 211–35.
- BAUMOL, W. J. (1967): *American Economic Review*, 57, 415–426.
- (2001): *Paradox of the services: exploding costs, persistent demand*, vol. 1 of *The growth of service industries : the paradox of exploding costs and persistent demand*, Edward Elgar, Cheltenham.
- BECKER, G. S. (1964): *Human capital: A theoretical and empirical analysis*, 11271, National Bureau of Economic Research.
- BLANCHARD, O. J. (1985): “Debt, Deficits, and Finite Horizons,” *Journal of Political Economy*, 93, 223–47.

- BLANCHARD, O. J. AND S. FISCHER (1989): *Lectures on Macroeconomics*, vol. 1 of *MIT Press Books*, The MIT Press.
- BUERA, F. J. AND J. P. KABOSKI (2009a): “Can Traditional Theories of Structural Change Fit The Data?” *Journal of the European Economic Association*, 7, 469–477.
- (2009b): “The Rise of the Service Economy,” .
- (2011): “Scale and the origins of structural change,” *Journal of Economic Theory*, 1, 1–29.
- CAMERON, S. V. AND J. J. HECKMAN (2001): “The Dynamics of Educational Attainment for Black, Hispanic, and White Males,” *Journal of Political Economy*, 109, 455–499.
- CARD, D. AND T. LEMIEUX (2001): “Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis,” *The Quarterly Journal of Economics*, 116, 705–746.
- CARDAK, B. A. (2004): “Education Choice, Endogenous Growth and Income Distribution,” *Economica*, 71, 57–81.
- CLARK, C. (1940): *The Conditions of Economic Progress*, London/New York.
- DE FRAJA, G. (2002): “The Design of Optimal Education Policies,” *Review of Economic Studies*, 69, 437–66.
- ECHEVARRIA, C. (1997): “Changes in Sectoral Composition Associated with Economic Growth,” *International Economic Review*, 38, 431–52.
- EICHER, T. (2001): “Inequality and growth: the dual role of human capital in development,” *Journal of Development Economics*, 66, 173–197.
- GALOR, O. AND O. MOAV (2004): “From Physical to Human Capital Accumulation: Inequality and the Process of Development,” *Review of Economic Studies*, 71, 1001–1026.

- GLOMM, G. AND B. RAVIKUMAR (1992): "Public versus Private Investment in Human Capital Endogenous Growth and Income Inequality," *Journal of Political Economy*, 100, 813–34.
- GOLDIN, C. AND L. F. KATZ (2007): "The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005," NBER Working Papers 12984, National Bureau of Economic Research, Inc.
- GREENAWAY, D. AND M. HAYNES (2003): "Funding Higher Education in The UK: The Role of Fees and Loans," *Economic Journal*, 113, F150–F166.
- JONES, C. I. (1995): "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103, 759–84.
- KANE, T. J. (1994): "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, 102, 878–911.
- KOHLBERG, E. (1976): "A model of economic growth with altruism between generations," *Journal of Economic Theory*, 13, 1–13.
- LEWIN, T. (2008): "College May Become Unaffordable for Most in U.S." *The New York Times*.
- LOURY, G. C. (1981): "Intergenerational Transfers and the Distribution of Earnings," *Econometrica*, 49, 843–67.
- MADDISON, A. (2005): "Measuring And Interpreting World Economic Performance 1500-2001," *Review of Income and Wealth*, 51, 1–35.
- NGAI, L. R. AND C. A. PISSARIDES (2007): "Structural Change in a Multisector Model of Growth," *American Economic Review*, 97, 429–443.
- SCHALTEGGER, C. A. AND C. GORGAS (2011): "The Evolution of Top Incomes in Switzerland over the 20th Century," CREMA Working Paper Series 2011-06, Center for Research in Economics, Management and the Arts (CREMA).

VEALL, M. (2010): “Top Income Shares in Canada: Updates and Extension,” *WORKING PAPER*.

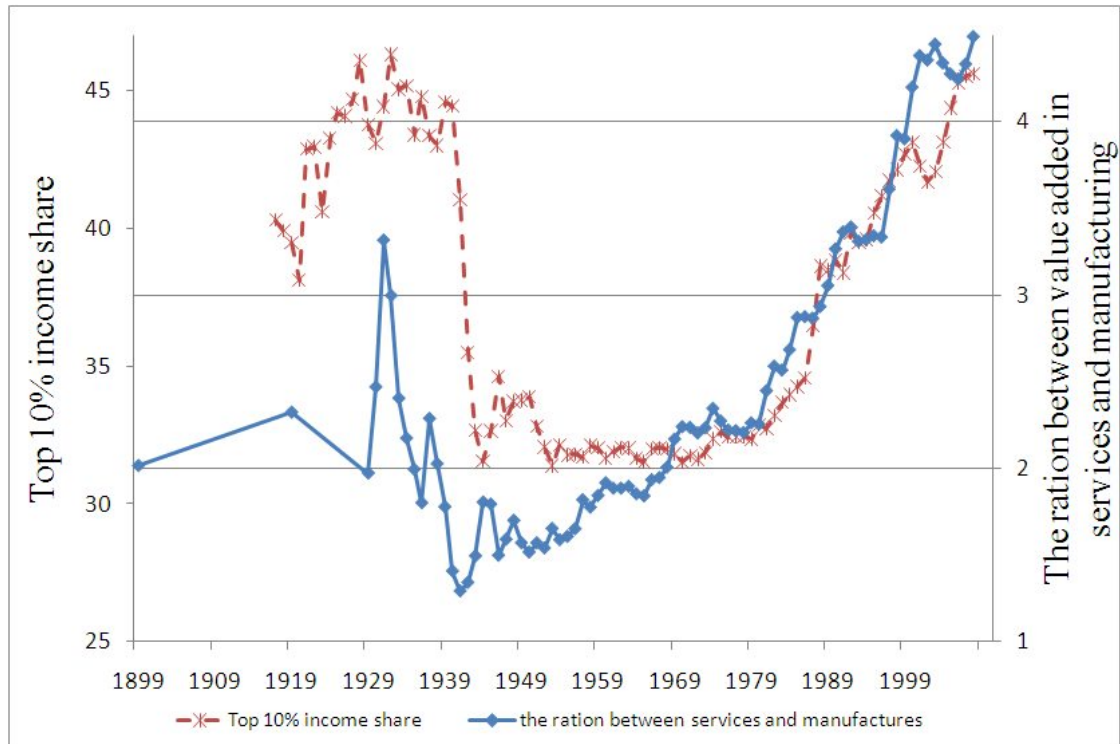
WALKER, I. AND Y. ZHU (2008): “The College Wage Premium and the Expansion of Higher Education in the UK,” *Scandinavian Journal of Economics*, 110, 695–709.

WEISS, M. (2008): “Skill-biased technological change: Is there hope for the unskilled?” *Economics Letters*, 100, 439–441.

YAARI, M. E. (1965): “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer,” *The Review of Economic Studies*, 32, 137–150.

2.7 Appendix

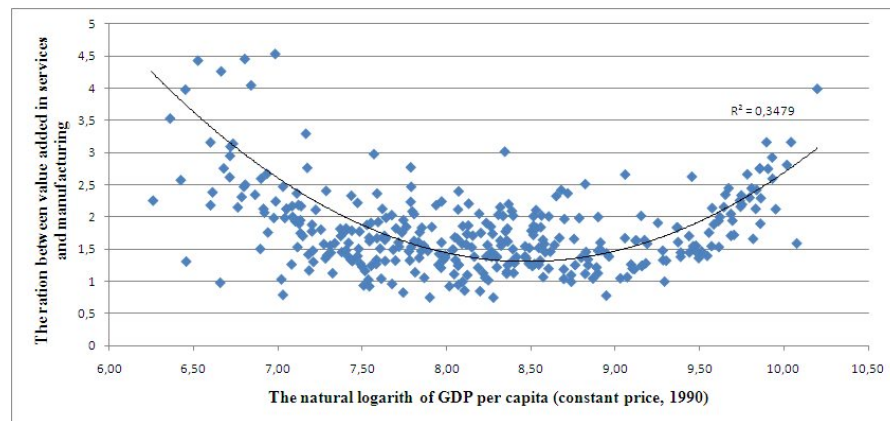
Figure 2.1: Top 10% income share and the relative share of services in the US



Sources: Author's calculations based on

1. The top 10% income share: Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty and Emmanuel Saez, The World Top Incomes Database, <http://g-mond.parisschoolofeconomics.eu/topincomes>;
2. The value added share of GDP by major sector (agriculture, industry, and services): National Income and Product Accounts (NIPA).

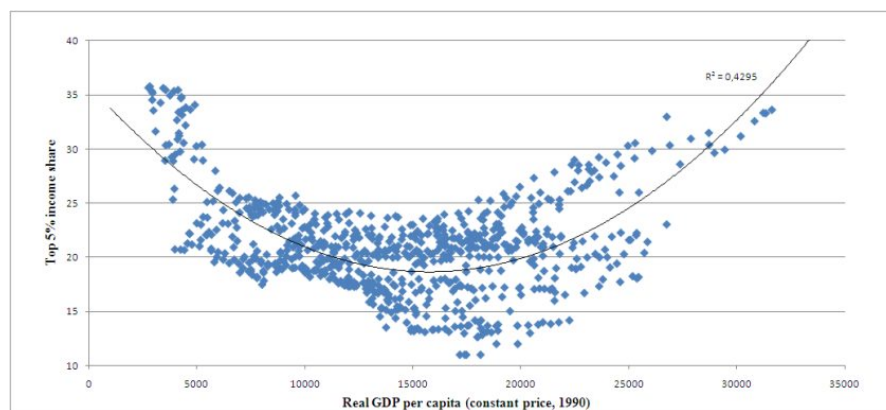
Figure 2.2: Ratio of Services to Manufacturing in value-added vs. Log of Income per capita (Country Panels)



Sources: Author's calculations based on

1. The value added share of GDP by major sector (agriculture, industry, and services) Buera and Kaboski (2011)
2. The long longitude real (1990 base year) income per capita data: Maddison (2005)

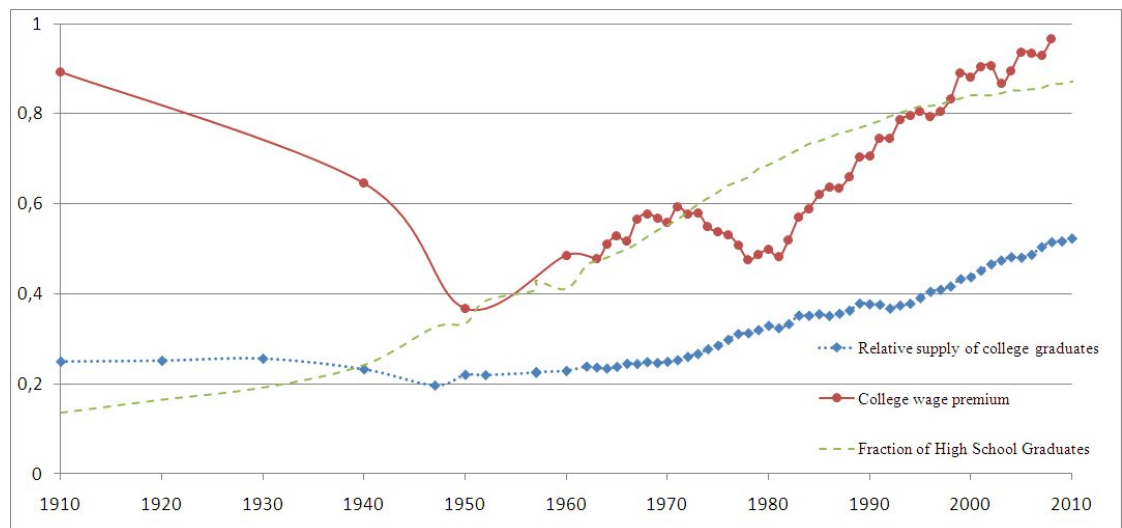
Figure 2.3: the Top 5% income share vs. GDP Per capita (Country Panels)



Sources: Author's calculations based on

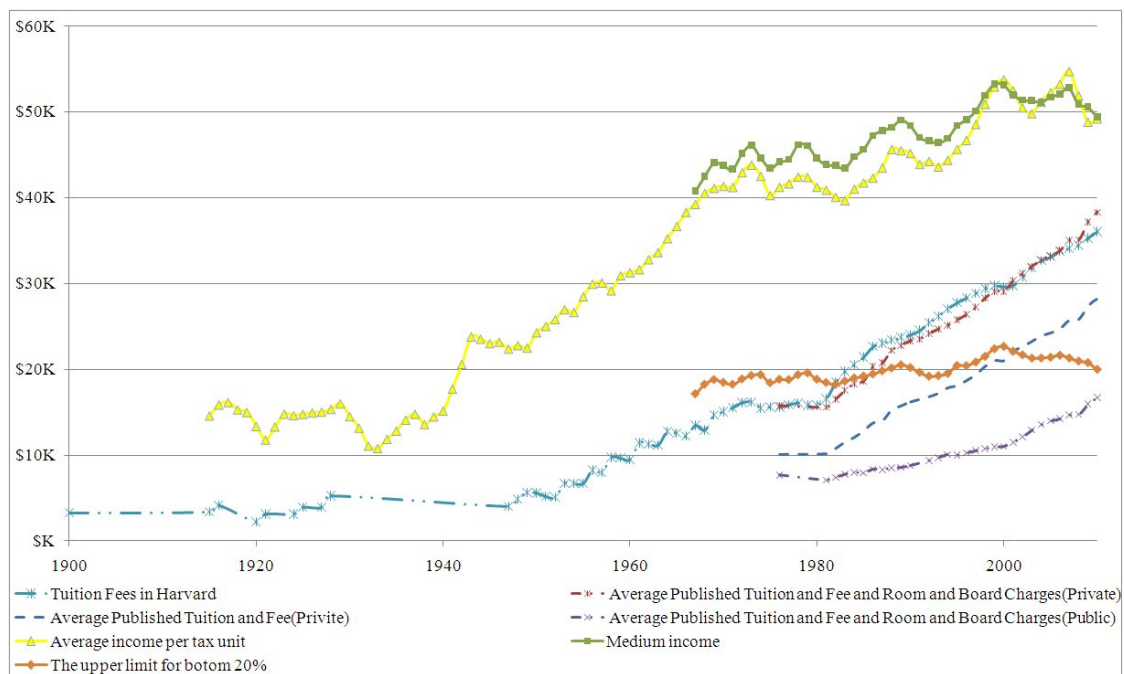
1. The top 5% income share: Alvaredo, Facundo, Anthony B. Atkinson, Thomas Piketty and Emmanuel Saez, The World Top Incomes Database, <http://g-mond.parisschoolofeconomics.eu/topincomes>; Canada 2000-2007: Veall (2010); Iceland: Inland Revenue; Switzerland 1997-2007: Schaltegger and Gorgas (2011).
2. The real (1990 base year) GDP per capita: Total Economy Database <http://www.conference-board.org/data/economydatabase/>

Figure 2.4: Relative supply of college skills and the college wage premium



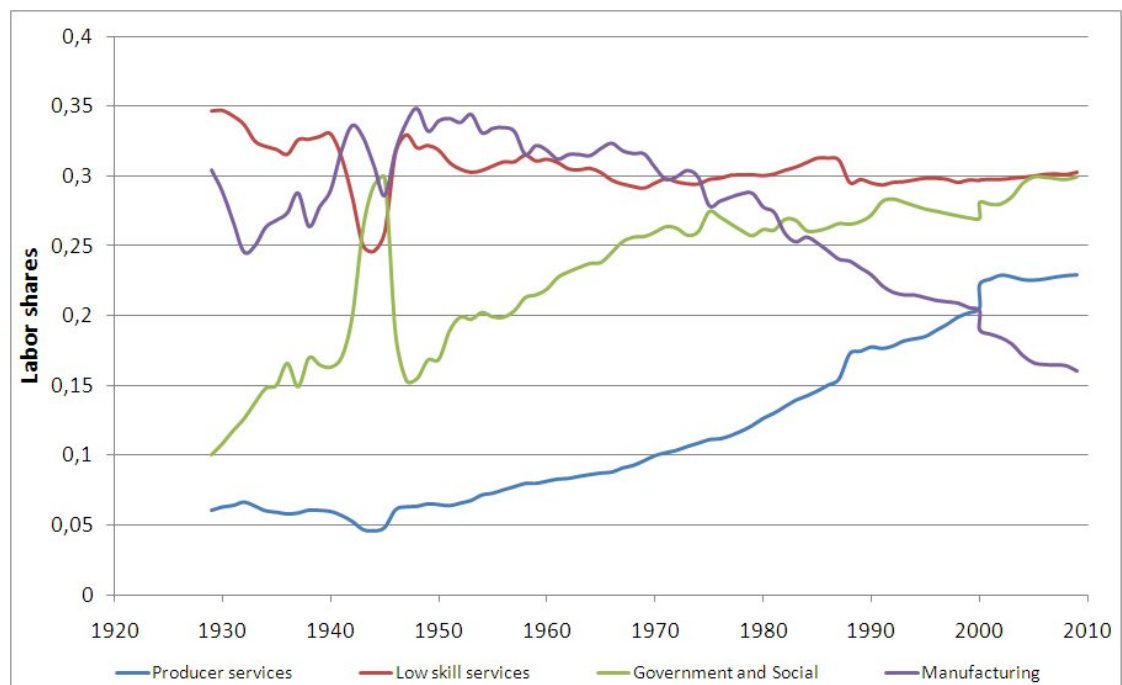
Sources: US Department of Commerce, Census Bureau; Goldin and Katz (2007);
Acemoglu and Autor (2011)

Figure 2.5: Tuition Fees vs. Median Income



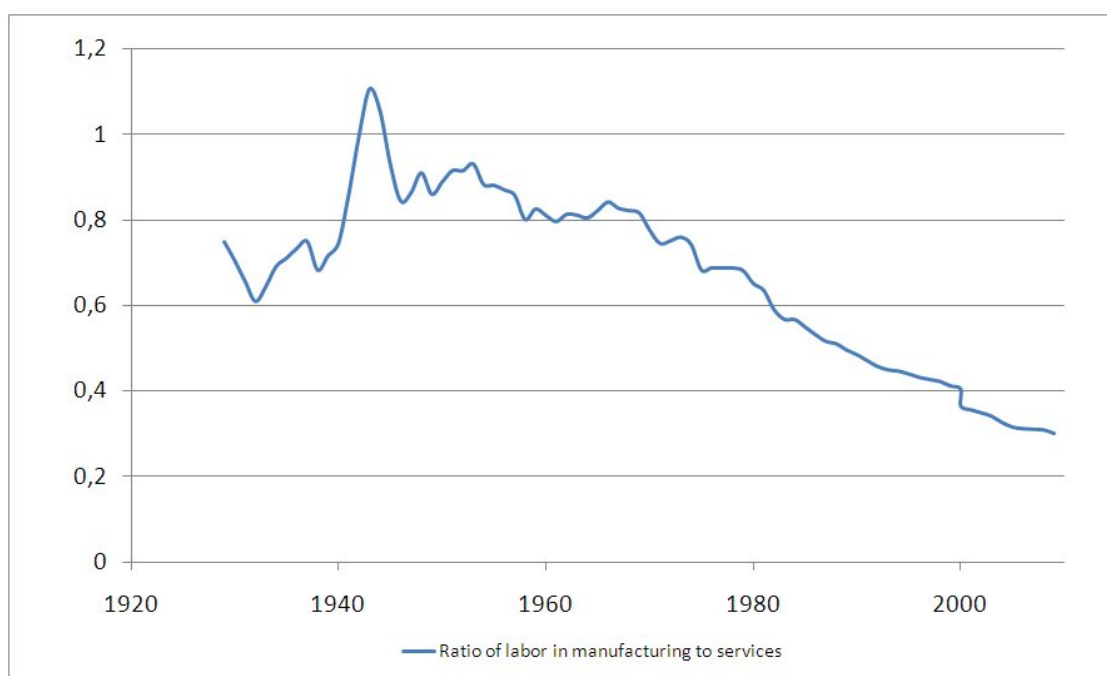
Sources: the Harvard Fact Book and the Harvard *Crimson*; Trends in College Pricing (2011).http://trends.collegeboard.org/college_pricing/; the World Top Incomes Database; the US Census Bureau.

Figure 2.6: The growth of Low and High skill industries shares: Labor



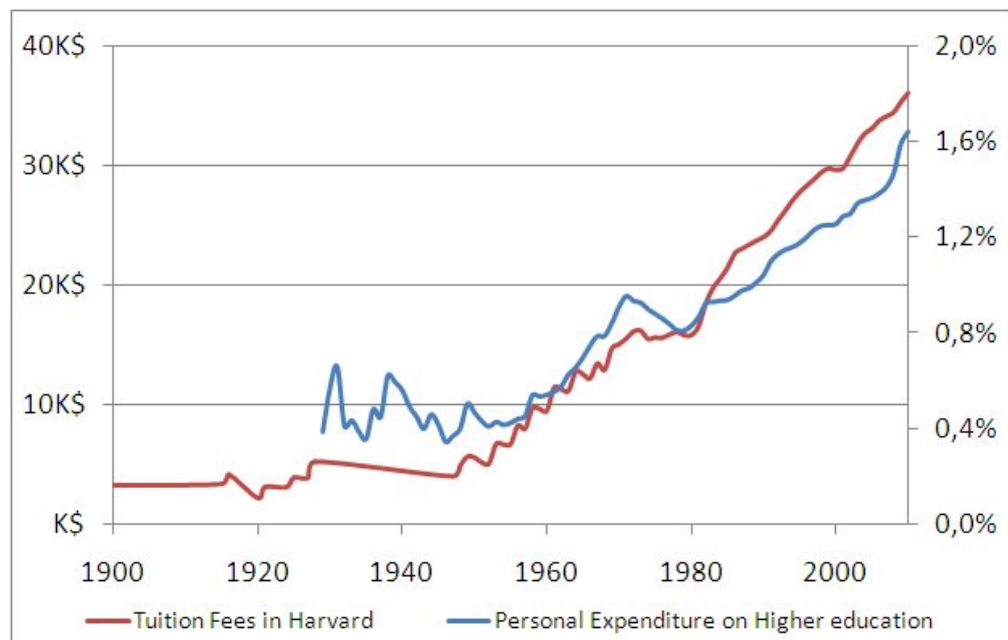
Source: National Income and Product Accounts (NIPA) data.

Figure 2.7: The ratio between labor shares in manufacturing and low-skill services



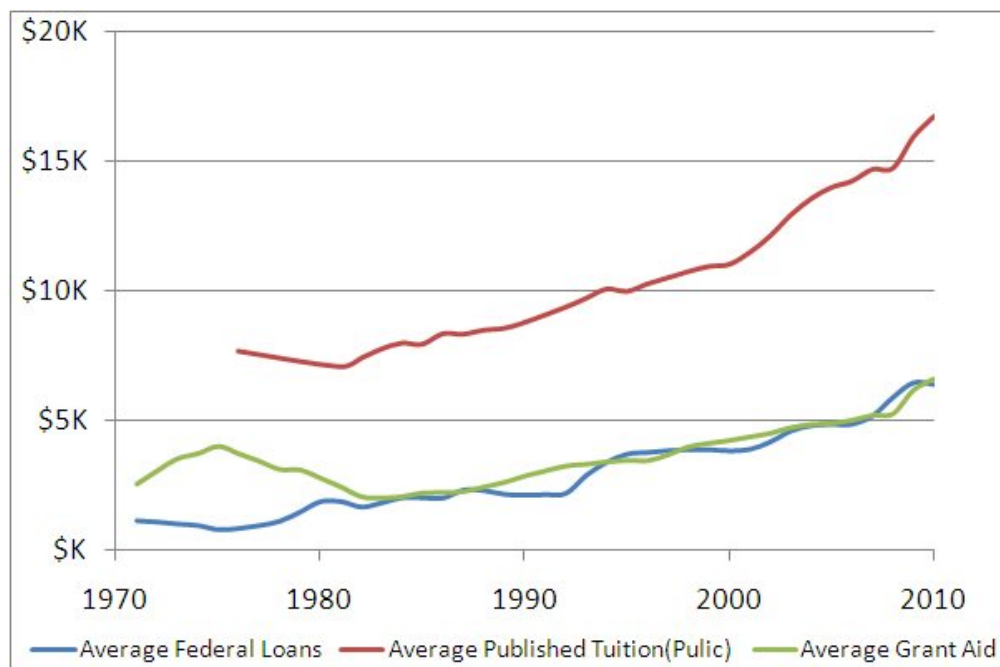
Source: National Income and Product Accounts (NIPA) data.

Figure 2.8: Tuition fees and expenditures on the higher education



Sources: the Harvard Fact Book and the Harvard *Crimson*; National Income and Product Accounts (NIPA) data.

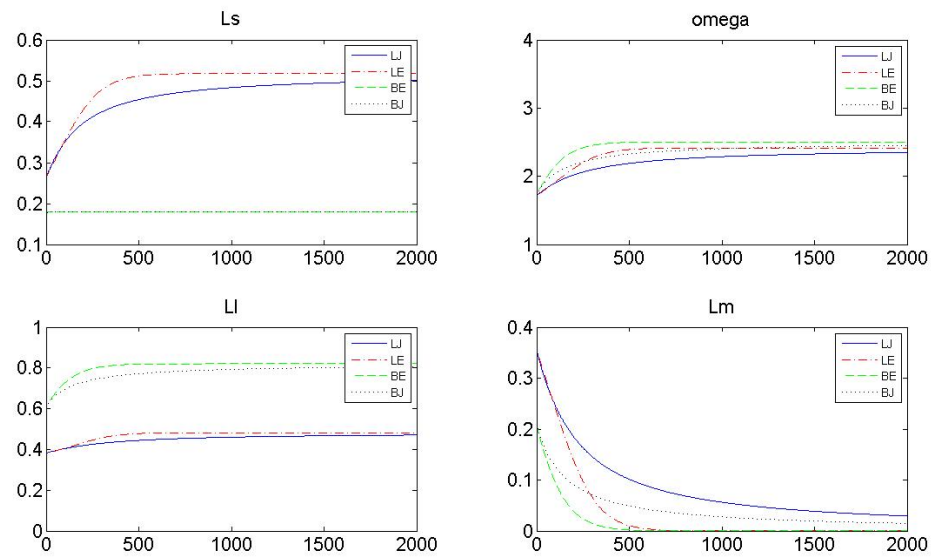
Figure 2.9: The increase of student aids: grants and loans against the growth of tuition fees



Source: Trends in Student Aid

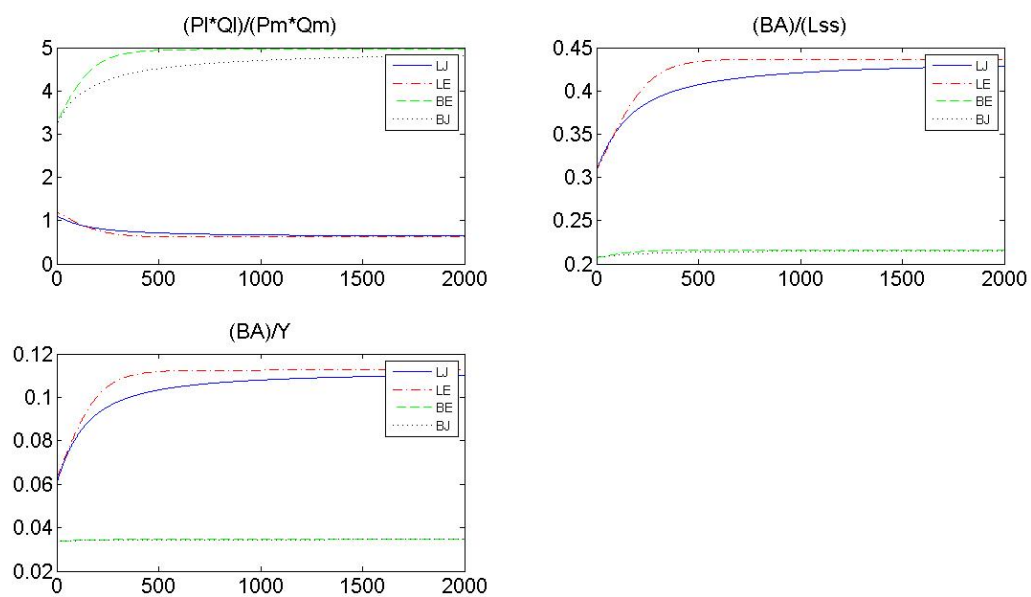
(2011).http://trends.collegeboard.org/college_pricing/.

Figure 2.10: Labor allocation and the college wage premium



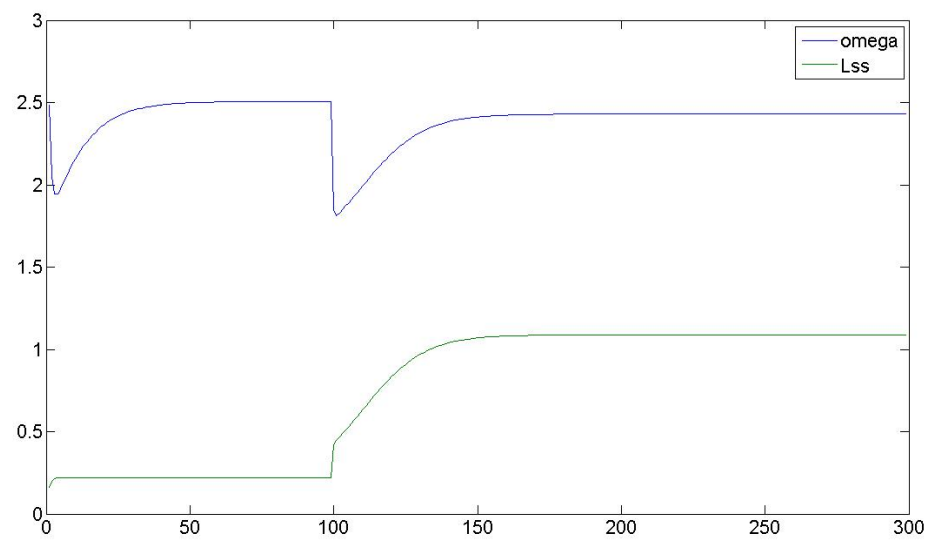
Source: Computer simulations.

Figure 2.11: The share of services and expenditure on higher education



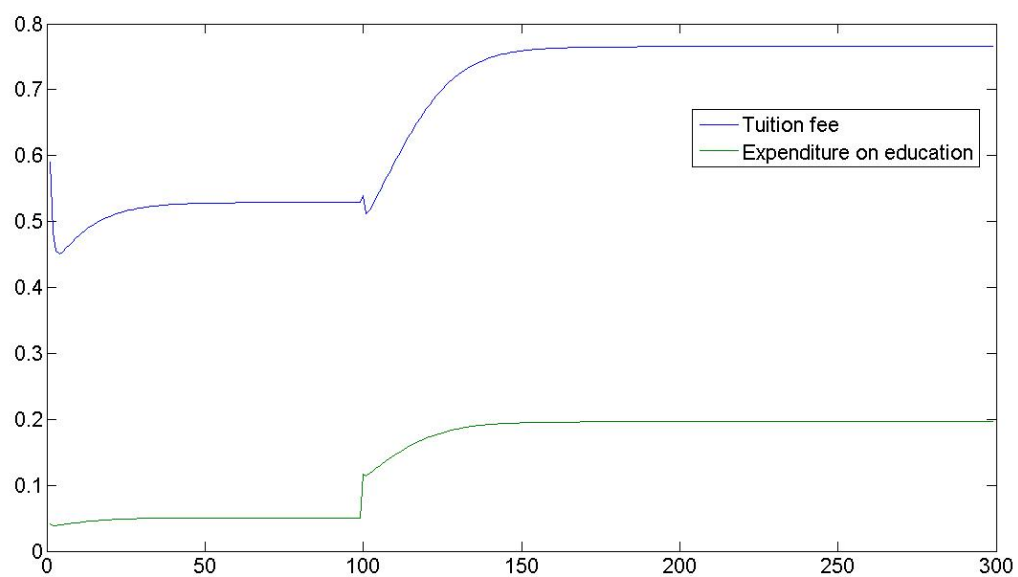
Source: Computer simulations.

Figure 2.12: Simulation: relative supply of college skills and the college wage premium



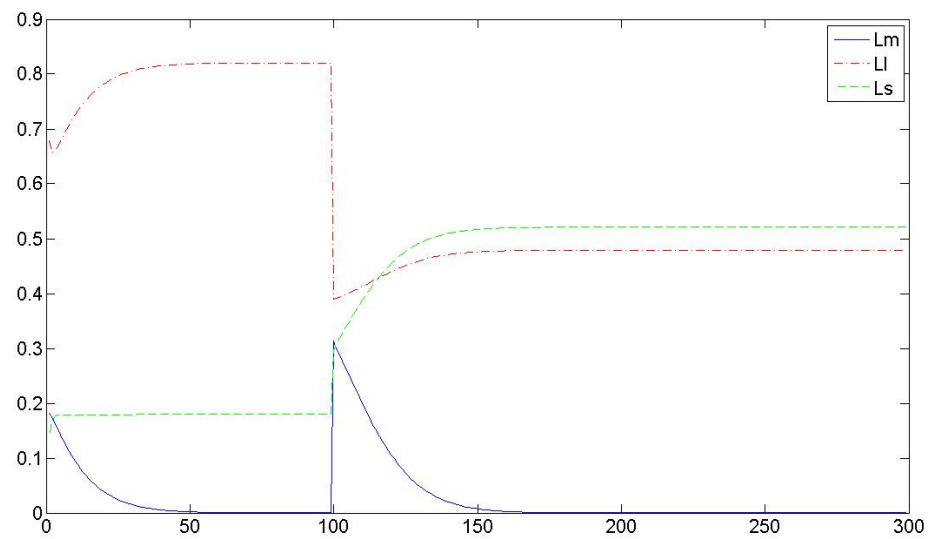
Source: Computer simulations.

Figure 2.13: Simulation: tuition fees and expenditures on higher education



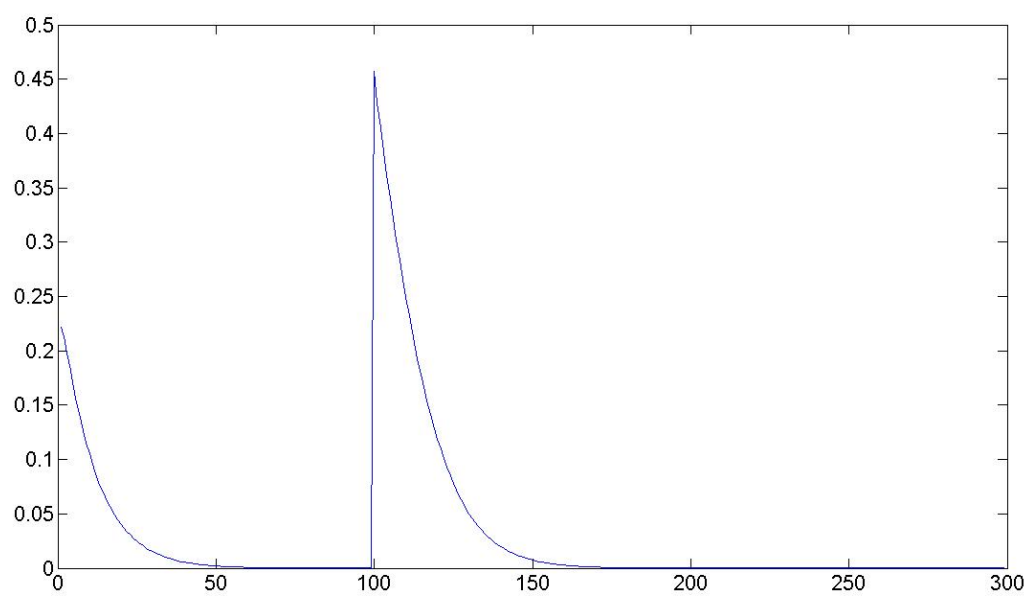
Source: Computer simulations.

Figure 2.14: Simulation: Labor allocation



Source: Computer simulations.

Figure 2.15: Simulation: the ratio between labor shares in manufacturing and low-skill services



Source: Computer simulations.

Table 2.1: High-skill and low-skill industries

id	industry	Fraction of college educated labor		Fraction of total man-hours		Fraction of value added		Compensation per employee	
		1940	2000	1940	2000	1947	2000	1940	2000
	All industry	14,7%	53,4%	100%	100%	100%	100%	1087	42888
	NON-SERVICES	6,0%	43,3%	48,3%	27,3%	41,4%	21,2%		
1	Mining	10,3%	32,6%	1,5%	0,5%	2,8%	1,2%	1409	69359
2	Agriculture and Fishing	1,8%	33,9%	22,0%	2,7%	8,0%	1,0%	150	15786
3	Construction	4,1%	35,2%	3,7%	7,5%	3,7%	4,4%	938	54771
4	Nondurable Drugs	9,3%	39,1%	10,9%	6,6%	13,3%	5,7%	1346	49934
5	Durable Office machines	28,5%	66,3%	0,0%	0,3%				
		9,7%	41,6%	10,3%	10,0%	13,5%	8,8%	1638	33157
		22,5%	65,7%	0,2%	0,4%				
	SERVICES	10,8%	59,3%	50,4%	71,7%	56,6%	79,6%		
	Low-skill	8,3%	43,0%	34,3%	28,0%	29,5%	24,4%		
6	Transportation Air transportation	7,2%	38,5%	5,1%	3,6%	5,7%	5,7%	1714	43281
		26,0%	65,1%	0,1%	0,5%				
7	Personal Services	5,6%	39,5%	8,7%	2,6%	2,8%	2,3%	678	16426
8	Retail Trade Household appliance	8,2%	41,0%	15,5%	14,7%	11,5%	6,7%	889	24012
		16,1%	65,9%	0,2%	0,8%				
9	Wholesale Trade Drugs	12,1%	47,6%	2,9%	3,8%	6,8%	6,0%	1666	54712
		22,3%	60,1%	0,0%	0,2%				
10	Utilities	15,5%	49,0%	1,3%	1,0%	1,5%	2,8%	1874	74676
11	Entertainment broadcasting	20,0%	57,3%	0,9%	2,2%	1,1%	0,9%	1025	32094
		31,7%	68,3%	0,1%	0,5%				

id	industry	Fraction of college educated labor		Fraction of total man-hours		Fraction of value added		Compensation per employee	
		1940	2000	1940	2000	1947	2000	1940	2000
	High-skill	16,1%	69,2%	17,4%	44,7%	28,7%	55,1%		
	<i>Production related services</i>	13,2%	66,9%	7,2%	16,4%	13,8%	36,0%		
12	Business and Repair Services	15,0%	64,7%	2,8%	8,9%	2,5%	13,5%	1497	40652
	Auto repair	5,7%	31,8%	1,2%	1,3%	0,4%			
	Misc repair	12,1%	40,4%	0,4%	0,5%	0,3%			
	Misc business	22,7%	66,2%	0,3%	3,9%	1,0%	4,7%		
	Advertising	23,6%	67,6%	0,2%	0,4%				
	Accounting	31,4%	81,3%	0,0%	0,6%				
	Legal services	25,7%	81,6%	0,8%	1,2%	0,4%	5,6%		
	Engineering	34,4%	82,2%	0,0%	1,0%	0,4%	3,2%		
13	Telecommunications	12,4%	65,2%	0,9%	1,1%	1,4%	2,8%	1802	75174
	Telephone	9,7%	64,0%	0,7%	0,8%				
	Telegraph	25,7%	68,0%	0,1%	0,3%				
14	FIRE	14,4%	70,1%	3,5%	6,4%	10,0%	19,7%	1609	64469
	Estate offices	12,4%	21,5%	0,0%	0,0%				
	Real estate	14,1%	63,0%	1,1%	1,5%	7,5%	12,1%		
	Banking	15,0%	68,7%	1,1%	2,3%	1,4%	3,4%		
	Insurance	14,3%	71,6%	1,3%	1,9%	1,0%	2,4%		
	Invest companies	27,4%	83,1%	0,0%	0,8%	0,1%	1,7%		
	<i>Government related services</i>	17,3%	70,5%	10,2%	28,4%	14,8%	19,2%		
15	Government	12,2%	67,5%	4,3%	6,7%	12,5%	12,3%	1410	52845
	State	16,5%	26,9%	0,0%	0,0%				
	Postal service	11,0%	57,3%	0,7%	0,7%				
	Federal	12,6%	67,5%	1,5%	3,4%	8,4%	3,9%		
	Local	12,4%	70,2%	2,1%	2,7%	4,1%	8,4%		
16	Social Services	21,0%	71,5%	5,9%	21,6%	2,4%	6,9%		
	Welfare and religious services	23,7%	62,0%	0,9%	1,1%				
	Membership organizations	23,7%	63,3%	0,0%	0,5%	0,5%			
	Health services	22,2%	65,2%	0,0%	4,7%		3,1%		
	Misc professional	36,0%	66,7%	0,0%	2,5%				
	Hospitals	15,8%	74,3%	2,4%	3,9%	1,5%	2,4%	610	43787
	Education	24,7%	76,6%	2,7%	8,9%	0,3%	0,8%	1084	31455

Source: Author's calculations based on Buera and Kaboski (2009b).

Table 2.2: Final vs. Intermediate consumption

	Final Household Consumption		Intermediate Consumption		Net Export		Other	
	2005	1995	2005	1995	2005	1995	2005	1995
All industry	37,6%	36,6%	46,2%	46,3%	-3,1%	-0,7%	19,2%	17,8%
Agriculture	16,9%	12,6%	85,5%	88,2%	-2,3%	2,1%	-0,1%	-2,9%
Mining	2,2%	1,6%	130,3%	123,3%	-54,4%	-26,6%	21,9%	1,6%
Food	65,9%	60,9%	37,0%	37,9%	-4,3%	1,0%	1,4%	0,2%
Textiles	163,6%	78,7%	61,9%	55,7%	-132,3%	-36,3%	6,9%	1,9%
Chemicals	36,4%	25,6%	72,5%	72,4%	-10,5%	0,9%	1,6%	1,1%
Metals	0,7%	0,4%	124,4%	107,8%	-27,6%	-8,9%	2,5%	0,7%
Vehicles	45,8%	40,7%	47,0%	34,8%	-25,4%	-19,6%	32,6%	44,1%
Utilities	44,1%	34,6%	55,6%	65,4%	-0,5%	0,0%	0,8%	0,0%
Construction	0,0%	0,0%	12,1%	36,9%	0,0%	0,3%	87,9%	62,9%
Producer services	17,1%	17,6%	41,2%	38,2%	2,3%	1,0%		
Telecommunications	29,8%	36,1%	65,9%	60,6%	1,2%	1,3%	3,1%	2,0%
Finance and Insurance	41,6%	51,2%	56,1%	46,1%	1,8%	2,8%	0,6%	0,0%
Real estate	59,4%	61,5%	32,0%	33,1%	0,0%	2,4%	8,5%	2,9%
Information	8,7%	2,3%	55,5%	91,8%	2,0%	2,9%	33,8%	3,0%
R&D	5,9%	0,5%	89,6%	98,0%	0,4%	1,4%	4,1%	0,1%
Business Activities	14,0%	10,6%	78,8%	86,9%	7,2%	0,7%	0,0%	1,7%
Social and Government								
Public admin.	15,1%	0,0%	8,0%	1,2%	0,2%	0,0%	76,7%	98,8%
Education	81,1%	25,9%	17,4%	4,2%	0,0%	0,0%	1,5%	69,9%
Health	97,7%	88,7%	2,1%	3,1%	0,0%	0,0%	0,2%	8,2%
Low skill								
Retail and Wholesale	59,5%	54,9%	30,1%	35,2%	4,7%	4,1%	5,7%	5,8%
Recreation	78,5%	75,6%	21,2%	24,3%	0,1%	0,2%	0,2%	-0,1%
Transportation	23,5%	23,9%	64,5%	62,5%	8,1%	11,0%	3,9%	2,5%
Personal services	39,1%	66,4%	60,0%	32,1%	0,5%	1,7%	0,4%	-0,2%

Source: Author's calculations based on OECD input-output database.

Chapter 3

Sovereign Debt Issuance and Selective Default

3.1 Introduction

“A national debt, if it is not excessive, will be to us a national blessing.”

— Alexander Hamilton

Humanity has witnessed sovereign debt crises for hundreds of years. The first recorded instance of sovereign default dates back to 377 B.C. in ancient Greece. Sovereign default has been studied extensively in the literature. However, the focus has mainly been on default on external debt, while the study of domestic defaults has been neglected. Reinhart and Rogoff (2011b) have documented and categorized all default events in the last 210 years. Based on their observations, there have been at least 58 *de jure* defaults on domestic public debt. This is certainly an underestimate, due to the difficulty of detecting *pure* domestic defaults.¹ Also, out of 267 defaults in this period, only 17 times did the government default simultaneously on both domestic and foreign debt.

¹For example, the large-scale 1989 *pure* domestic default is relatively unknown outside Argentina. The most well know domestic default happened in Russia 1998, which was one of the largest local currency debt defaults (US \$39 billion). This number does not include *de facto* default through inflation, the nationalization of pensions and other forms.

In this paper, we address the open question of why governments usually default *selectively* on either foreign or domestic debt. We do so by providing a novel theory of domestic debt and default, where domestic debt is issued to smooth tax distortions, and combining it with the well established literature on foreign debt and default. We present a two-period model to deliver the economic intuition, and a calibrated quantitative model to replicate data moments. We show that our theory is empirically plausible, as it is able to match frequencies of different types of default and debt compositions. To the best of our knowledge, this is the first contribution that is able to replicate two stylized facts: that defaults happen mostly in a selective fashion, and that the composition of bondholders matters for interest rates and the volume of total public debt. The two-period version of the model is the starting point for an additional discussion of the role of secondary markets in solving sovereign default problems. Our analysis questions the efficiency role associated with secondary markets.

We build an incomplete markets model in which the government has limited commitment. The government has to cover its expenditures and has three means of financing them: it issues one-period defaultable bonds on an international and a domestic market, and it collects taxes. Tax collection is costly because taxes are distortionary. The economy is subject to two shocks: an output shock and a tax distortion shock. While the output shock provides incentives for the government to borrow on international markets, the tax distortion shock creates a wedge between domestic borrowing and taxation. This breaks Ricardian equivalence in our endowment economy, and draws a distinction between tax-financed and debt-financed expenditure policies. In this we provide a simple theory of domestic debt issuance.

Foreign debt can be used to smooth out both shocks, which makes it a more valuable instrument. However, if the correlation of the two processes is sufficiently low, then using only one instrument to smooth two shocks would result in households' consumption being too volatile. Therefore, the government engages in borrowing on both markets. When the government has outstanding debts on both markets it faces two trade-offs: one between foreign repayment and default, and

another between domestic repayment and default.

The mechanism of foreign default is similar to that in Eaton and Gersovitz (1981) and Arellano (2008). A benevolent government accumulates defaultable foreign debt in order to smooth residents' consumption over the business cycle. Interest rates reflect default probabilities, which are endogenous to the borrower's incentives to default. The government decides each period whether to transfer resources away from the economy as a repayment of debt to foreign investors or to keep resources at home and suffer default penalties. When output is low, *ceteris paribus*, it is more costly for a risk averse borrower to respect the contract. Default occurs along the equilibrium path after a long enough sequence of negative output shocks. These contributions gave rise to a large literature, which has nonetheless not yet considered domestic debt and default in an open economy setting.

While the mechanisms and trade-offs behind foreign default are clear, the domestic default literature is still at an early stage. There are two recent contributions that adhere to the benevolent government assumption and study domestic default in a Ramsey setting. D'Erasmus and Mendoza (2013) propose a heterogeneous agent model in which a utilitarian government relies on lump-sum taxes and defaultable bonds to finance stochastic governments expenditures. Default has a redistributive aspect, because it hurts mostly the rich, while repayment by taxation hurts mostly the poor. Pouzo and Presno (2014), on the other hand, consider a model in which the government relies on distortionary labor income taxes and defaultable bonds to finance its stochastic expenditures. The government might default to mitigate these distortions. The second crucial trade-off in our model, the one behind domestic debt and default, is similar to their mechanism. Both contributions, however, are closed-economy models that do not consider borrowing on international markets.

Both repayment and default on domestic debt are transfers of resources within the economy. In a case of default on domestic debt, the government suffers default penalties similar to the penalties imposed after foreign default. When the government decides to repay, it needs to finance this repayment by collecting taxes. When distortions from taxation are high, the government prefers to issue debt

rather than collect taxes, hoping that in the future tax collection will be less distortionary, giving it the ability to repay the debt at a lower cost. The government thus issues domestic debt up to an endogenous debt limit, and if the possibility of repayment through non-distortionary taxes does not arrive it has no other choice than to default.

Vasishtha (2010) and Erce (2012) study the selective nature of sovereign default with foreign and domestic investors. The former generates domestic debt issuance through disutility of taxation, but in equilibrium foreign default never happens. In the latter, both domestic and foreign debt levels are exogenously predetermined. Our analysis shows that incorporating two shocks, to output and to taxation, is crucial to generating equilibria with both types of selective default, and that the feedback loop from selective default to debt issuance should not be neglected. In our paper, both domestic and foreign debt issuance and selective default are optimal decisions of the government. In addition, Cooper et al. (2008) study how the distribution of debt among domestic and foreign investors influences the government's incentives to default. They find conditions (government expenditure and the fraction of debt held by foreign investors being high enough) under which the government has incentives to default, but the underlying composition of debt is given exogenously. In this paper, we derive endogenous fractions of public debt held by domestic and foreign investors.

The main contributions of this paper are the new theory of selective sovereign defaults and a quantitative framework to study sovereign debt issuance and debt composition. But our analysis also has also some other, quite novel implications. After the Great Recession, secondary sovereign debt markets attracted increasing interest among economists. Based on the two-period version of our model, we analyze the role of secondary markets in solving the problem of sovereign risk. Broner et al. (2010) show that, even in the absence of default penalties, sovereign risk does not prevent governments from borrowing on international markets if foreign creditors can resell their assets to domestic investors on secondary markets. We show that the key assumption behind their result is that tax collection is costless. We show conditions for which their result does not survive with costly

taxation (the supply of defaultable bonds is high compared to demand). The result of the trade depends on how well each group of investors can coordinate their actions. In particular, without any coordination, trade on secondary markets generates a possible welfare loss, as it incentivizes the government to default on all its debt, instead of only foreign debt. We also prove that whenever secondary markets fail to reduce the default problem, debt haircuts can play a useful role, and *vice versa*.

The remainder of the paper is organized as follows. In the next section we summarize empirical facts on domestic and foreign public debt holdings and selective defaults. Section 3 studies equilibrium in a two-period model and shows intuitively the main trade-offs. Section 4 presents an infinite-horizon version of the model and the results of a calibration exercise. Section 5 analyzes the role of secondary markets and haircuts. The last section concludes.

3.2 Facts

The goal of this section is to establish three stylized facts that motivate our analysis. First, that sovereign defaults happen mostly in a selective fashion; second, that governments have a number of tools to discriminate among different types of bondholders; and third, that the composition of bondholders matters. In this section we review some empirical studies of selective sovereign defaults and the composition of bondholders, and augment them with our findings.

Before we begin our discussion, we set the scene with some definitions. There are three different ways to draw the distinction between domestic and foreign debt. According to the *legal* definition, domestic debt is any debt issued according to domestic law, regardless of its currency, and regardless of who holds it. According to the *economic* definition, domestic debt is held by residents, regardless of the currency and the law under which it was issued. Finally, according to the *currency* definition, domestic debt is the debt denominated in home currency, regardless of law and the residency of bond holders. The second definition creates clear differential incentives for the sovereign to default. For this reason, throughout the

model, we adopt the *economic* definition.

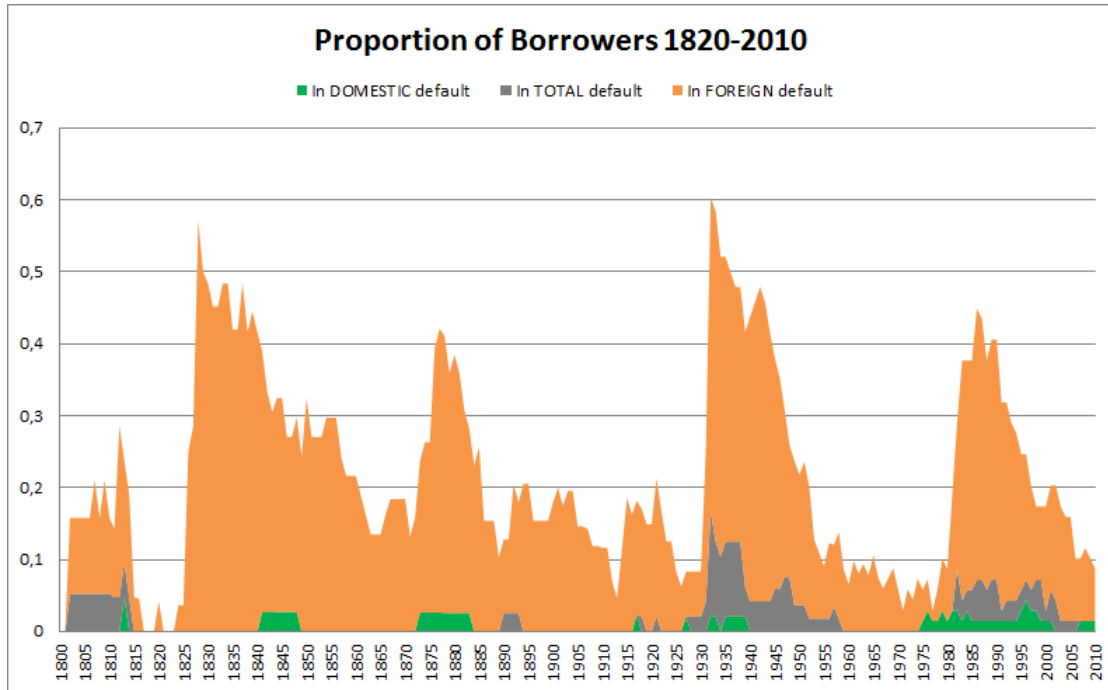
An important point to raise is that these three definitions do not necessarily overlap. However, Reinhart and Rogoff (2011a) claim that: “*The overwhelming majority of external public debt, debt under the legal jurisdiction of foreign governments, has been denominated in foreign currency and held by foreign residents*”. This was certainly true before the wave of capital flow liberalizations starting in the 1980s. After this, the mapping between the *legal* and the *economic* definitions is less ideal.² Still, we observe selective sovereign defaults both before and after the wave of capital flow liberalizations.

For our stylized facts and in our calibration we rely on three sources of data. Merler and Pisani-Ferry (2012) provide the breakdown of the public debt by the residence of holders for ten industrialized economies between 1990 and 2012. For the developing economies we rely on the dataset compiled by Panizza (2008), which covers the data of up to 130 countries between 1990 and 2007. Data for developing economies is however obtained using the *legal* definition. Our third source is the dataset on crises and defaults provided by Reinhart and Rogoff (2011b), which covers up to 70 countries between 1800 and 2010. The *legal* definition of debt is also used for the default data. In what follows we present three empirical facts that motivate and guide our theoretical analysis.

1. Sovereign defaults usually happen in a selective fashion. The database collected by Reinhart and Rogoff (2011b) reveals interesting features of sovereign default episodes between 1800 and 2010. First, domestic debt, usually neglected in the theoretical literature on sovereign risk, plays an important role in the build-up, during and after sovereign defaults on foreign holdings. This argument is

²A notable example here is the Mexican crisis of 1994. Short-term securities called *tesobonos* were dollar-denominated (foreign debt according to the *currency* definition), issued according to Mexican law (domestic debt according to the *legal* definition) and held by investors both in the US and Mexico (partly domestic and partly foreign debt according to the *economic* definition). Also, at that time, there were no means of tracing the final creditor either by nationality or by residence. Therefore, a default on *tesobonos* obligations could not have been classified either as domestic or as foreign selective default. Luckily, Mexico did not default in 1994.

Figure 3.1: Fraction of countries in different types of default



Source: Own calculations based on Reinhart and Rogoff (2011b)

extensively developed in Reinhart and Rogoff (2011a). Second, sovereign defaults happen on both domestic and foreign debt holdings, usually in a selective fashion. Whereas foreign defaults are common, domestic defaults are hardly rare. Out of 267 episodes of sovereign debt crisis identified across 70 countries in the last 210 years, 205 were *pure foreign*, 26 were *pure domestic*, and 36 featured government default on *both domestic and foreign* debt. Only 17 times did the default on home and foreign debt happen within the same year. In Figure 3.1 we plot the fraction of sovereign borrowers that were in foreign, domestic or total default in a given year between 1800 and 2010. These findings suggest that the assumption that sovereigns can default selectively fits reality better than the two alternative assumptions commonly used in the literature: that domestic debt is always senior and so only foreign debt is defaulted on, or that defaults can only be non-discriminatory.

2. Governments have a number of tools to discriminate among types of bondholders. How can the government default on foreign investors while repaying domestic investors or *vice versa*? The assumption that the two types of bondholders are indistinguishable, and therefore sovereigns can only default on total debt outstanding, underestimates the creativity of governments.

Among the tools that governments use to discriminate against particular types of bondholders, the most popular are capital controls, exchange controls and freezes on deposits. In 1990 Brazil defaulted on its domestic debt but kept servicing its foreign debt. All foreign exchange transactions were directed through the central bank and a multiple exchange rate regime was introduced as well as a freeze on local currency deposits.

In 1998 Russia defaulted on both foreign and local currency debt, imposing capital and exchange rate controls. However, in subsequent years Russia “undeclared” on its foreign obligations and kept servicing debts to foreign investors. Moreover, bonds held by domestic companies were also repaid, so Russia effectively defaulted only on domestic households’ holdings of public debt. Default was accompanied by both foreign and local currency deposit freezes.

Argentina’s 2001 default is often considered as a model case of foreign default.³ In fact, this episode is cataloged as a total default. First, all resident-held bonds, both domestic and foreign currency denominated, were converted to government-guaranteed loans, which were all later converted to pesos at a much lower rate than the market exchange rate. Also, 60% of the debt defaulted on in December 2001 was held by Argentines.

Recent examples of what could be considered pure foreign default (in peaceful times) include: Bolivia in 1989 (most domestic debt was repurchased a year before default), Pakistan in 1999 (which stopped payments on outstanding obligations to creditors in the UK, Europe and the US and put a freeze on foreign currency deposits mostly owned by non-residents) and most probably Cyprus in

³Many sovereign default models are calibrated to mimic salient features of this default (e.g. Arellano (2008)).

2013 (freeze and partial expropriation of deposits exceeding €100,000, which were mostly owned by non-residents).

3. The composition of bondholders matters. Empirical work on the composition of bondholders is growing. We draw on this literature, particularly on Andritzky (2012) and Dell’Erba et al. (2013), to show that the composition of investors is correlated with interest rates and the total level of debt to GDP. Dell’Erba et al. (2013) find that there is a significant correlation between spreads and debt levels when the majority of the debt is denominated in foreign currency (in both emerging economies and Eurozone countries). They also document that financial crises have more profound effects on economies that rely more on foreign borrowing. Andritzky (2012) finds a strong positive correlation between the fraction of domestic debt in total debt and the total debt-to-GDP ratio, and a negative correlation between the fraction of foreign debt and spreads in advanced economies. The present paper contributes to this literature by providing a framework to study the driving forces behind debt composition and its consequences for spreads, total debt and default incentives.

3.3 Two-period model

We begin by introducing the model in a simplified and tractable two-period version. We study an endowment economy that consists of three types of actors: domestic households, foreign investors and a benevolent government. The government can raise resources in three different ways: by issuing bonds to domestic households, by issuing bonds to foreign investors and by collecting taxes. Taxes are lump sum, but collecting taxes comes at a cost to the economy. We assume raising an amount T of taxes by the government induces a loss of $T(1 + \tau)$ resources to agents. This is a key element that will break Ricardian equivalence in this endowment economy and create a trade-off between taxes and domestic debt.

Domestic households are identical and risk averse. The representative household decides on her bond holdings to maximize lifetime utility subject to two

intra-period budget constraints:

$$\max_{b^h} (u(c_1) + \beta \mathbb{E}[u(c_2)]) \quad (3.1)$$

subject to:

$$y_1 = c_1 + T_1(1 + \tau_1) + q_h b_h, \quad (3.2)$$

$$y_2 + (1 - d_h)b_h = c_2 + T_2(1 + \tau_2), \quad (3.3)$$

where y is the exogenous output, c is consumption, T is taxes, τ is the distortion imposed by taxes, b_h is domestic bond holdings, q^h is the discount price of domestic bonds and d_h is the government's decision to repay ($d_h = 0$) or default on ($d_h = 1$) domestic debt.

Foreign investors are risk neutral and have deep pockets. They borrow on international markets at risk-free rate r and lend funds to the government at discount q_f to break even in expectation:

$$q_f = \frac{\mathbb{E}[1 - d_f]}{1 + r} \quad (3.4)$$

The government has to cover expenditures only in the first period $g_1 > 0$. Government expenditures in the second period are $g_2 = 0$. This creates an incentive to borrow due to the consumption smoothing motive. In the first period, the government decides on debt issuances in the domestic and foreign markets b_h and b_f . In the second period, the government takes repayment decisions d_h and d_f . The government maximizes the lifetime utility of domestic households subject to two intra-period government budget constraints:

$$g_1 = q_h b_h + q_f b_f + T_1, \quad (3.5)$$

$$(1 - d_h)b_h + (1 - d_f)q_f = T_2. \quad (3.6)$$

If the government decides to default, the economy will suffer proportional output penalties. After domestic default, output in the second period is reduced to

$$y_{hd} = y_2(1 - \delta_h), \quad (3.7)$$

and after foreign default, output in the second period is reduced to

$$y_{fd} = y_2(1 - \delta_f). \quad (3.8)$$

If the government decides to default on both markets, the economy will suffer from both output penalties.

Finally, in the second period the economy is subject to two shocks: an output shock and a tax distortion shock. Both processes are stochastic Markovian and assume two outcomes:

$$y_2 = \begin{cases} y_H & \text{with prob. } \Pi_y \\ y_L & \text{with prob. } 1 - \Pi_y, \end{cases} \quad (3.9)$$

$$\tau_2 = \begin{cases} \tau_L & \text{with prob. } \Pi_\tau \\ \tau_H & \text{with prob. } 1 - \Pi_\tau, \end{cases} \quad (3.10)$$

where subscript H stands for high and L for low.

If the debts are repaid with taxes, the government imposes distortions on the economy. If they are repaid with new debt, the government might go into default. The main driving forces of the government's optimal policies are two trade-offs. The first is the trade-off between a transfer of resources away from the economy as foreign debt repayment versus a loss of resources due to foreign default penalties. The second is between imposing distortions on the economy from tax collection versus imposing a loss of resources from domestic default penalties. Unlike in cases where Ricardian equivalence holds, the timing of taxes matters here.

3.3.1 Default schedule

We solve the model by backward induction starting in the second period. Given debt issuance decisions from the first period b_h and b_f , in the second period the government takes default decision that maximize domestic households' utility from consumption. As it is the terminal period there is no demand for government bonds in the second period, so the only source of income for the government is taxation. In the second period, four scenarios may arise: repayment, foreign default, domestic default and total default. Substituting the government's repayment decisions ($d_h \in \{0, 1\}, d_f \in \{0, 1\}$) and default penalties (3.7), (3.8) into households' second-period budget constraint (3.3) and the government's second-period

budget constraint (3.6), household consumption levels in each of the four scenarios are given by the following equations: (Notice that, in order to repay an amount b_h of domestic bonds to households, the government needs to raise $b_h(1 + \tau)$ taxes, which yields a net loss of τb_h resources to the economy.)

$$c^r = y_2 - b_f(1 + \tau_2) - b_h\tau_2, \quad (3.11)$$

$$c^{fd} = y_2(1 - \delta_f) - b_h\tau_2, \quad (3.12)$$

$$c^{hd} = y_2(1 - \delta_h) - b_f(1 + \tau_2), \quad (3.13)$$

$$c^{td} = y_2(1 - \delta_h)(1 - \delta_f), \quad (3.14)$$

where consumption superscripts stand for *repayment*, *foreign default*, *home default* and *total default* respectively.

A. Foreign default schedule

When deciding whether to default on foreign investors, the government compares household consumption under repayment and under foreign default. It is immediate to see that foreign debt will be repaid whenever:

$$\frac{b_f}{y} \leq \frac{\delta_f}{1 + \tau_2}, \quad (3.15)$$

where the left-hand side is the foreign debt-to-GDP ratio and the right-hand side is a number defined by parameters of the model. Whenever the inequality has the opposite sign, the government defaults on foreign debt.

Proposition 4. *If taxation is costly then the government's optimal policy on the international market is characterized by the foreign default threshold (3.15). Whenever the debt is below this threshold, it is riskless and is always repaid. Whenever it is above the threshold, it will always be defaulted on and therefore can never be issued. If either output or tax distortions are stochastic, the default threshold is also stochastic, debt can be risky and default can arise in equilibrium.*

Proof. The first part follows directly from comparing (3.11) and (3.12). For the second part, suppose that future output y_2 and tax distortions τ_2 are known in period one. Any debt b_f exceeding $\frac{y_2\delta_f}{1+\tau_2}$ will be defaulted on with certainty in

period two, therefore its discount price in period one is zero. The government is only able to take out loans $b_f \leq \frac{y_2 \delta_f}{1 + \tau_2}$ which are repaid with certainty. Foreign default cannot arise in equilibrium. For b_f to be in the default area with positive probability, we need at least one parameter to be stochastic. \square

B. Domestic default schedule

Similarly, we can define the domestic debt limit. Domestic debt will be repaid whenever:

$$\frac{b_h}{y} \leq \frac{\delta^h}{\tau_2}, \quad (3.16)$$

where the left-hand side is the domestic debt-to-GDP ratio and the right-hand side is a number defined by parameters of the model. Whenever the inequality has the opposite sign, the government defaults on domestic debt. Most importantly, the denominator on the right-hand side of inequality (3.16) is of a different magnitude than that in (3.15). This is because repayment of foreign debt is a transfer of resources out from the economy, while repayment of domestic debt is only a redistribution of resources within the economy. This redistribution is costly, and these costs are captured by the parameter τ_2 . Inequality (3.16) allows us to prove two interesting propositions.

Proposition 5. *If taxation is costless and home default induces small positive costs to the economy, then any level of domestic debt is repaid.*

This is the result of Broner et al. (2010), where taxes are assumed to be lump sum and default on domestic agents induces redistribution costs, which are endogenously derived (here captured by the parameter δ_h). This result has powerful consequences. For example, if any level of debt is sustainable on the domestic market, then if secondary debt markets are efficient, any level of foreign debt is also sustainable in repayment equilibrium. Foreign debt can always be repaid even without exogenous default penalties, and a sufficient solution to the default problem is to improve the efficiency of secondary debt markets.

Proposition 5 shows that the assumption of lump-sum taxes is the key to deriving the Broner et al. (2010) result. Without this assumption, there is finite limit to the amount of domestic debt that can be sustained in repayment equilibrium.

Proposition 6. *If taxation is costly then the government's optimal policy on the domestic market is characterized by the domestic default threshold (3.16). Whenever the debt is below this threshold, it is riskless and is always repaid. Whenever it is above the threshold, it will always be defaulted on and therefore can never be issued. If either output or tax distortions are stochastic, the default threshold is also stochastic, debt can be risky and default can arise in equilibrium.*

Proof. The first part follows directly from comparing (3.11) and (3.13). The proof of the second part is analogous to the proof of Proposition 4. \square

Inequalities (3.15) and (3.16) completely characterize government policy in the second period. Notice that whenever both inequalities are reversed, it is also the case that $c^{td} > c^r$, which is consistent with the definition of total default being simultaneous default on both domestic and foreign debts outstanding.

C. Default policies in the second period

Having established default thresholds in the second period, we posit an equilibrium in which, depending on the realizations of stochastic shocks, all four outcomes (repayment, foreign default, domestic default and total default) arise in the second period. The purpose of this part is to find a set of parameters that can sustain this equilibrium and, in the next subsection, to check that this set of parameters delivers debt issuances that are consistent with the posited equilibrium. By doing this we want to understand the mechanics and interactions between debt issuances and selective default, and prove that the set of parameters that is able to deliver the four outcomes is non-empty. Both of the stochastic processes in this economy have two outcomes. Therefore, we impose equilibrium conditions that would map the four possible realizations of joint (y, τ) stochastic processes into four equilibrium outcomes. These conditions are:

1. After a bad output shock $y_2 = y_L$, the government defaults on foreign debt regardless of the realization of the tax distortion shock.
2. After a good output shock $y_2 = y_H$, the government repays foreign debt regardless of the realization of the tax distortion shock.

3. After a bad tax distortion shock $\tau_2 = \tau_H$, the government defaults on domestic debt regardless of the realization of the output shock.
4. After a good tax distortion shock $\tau_2 = \tau_L$, the government repays domestic debt regardless of the realization of the output shock.

Mathematically these conditions can be summarized by four inequalities that follow from substituting realizations of y and τ into (3.15) and (3.16):

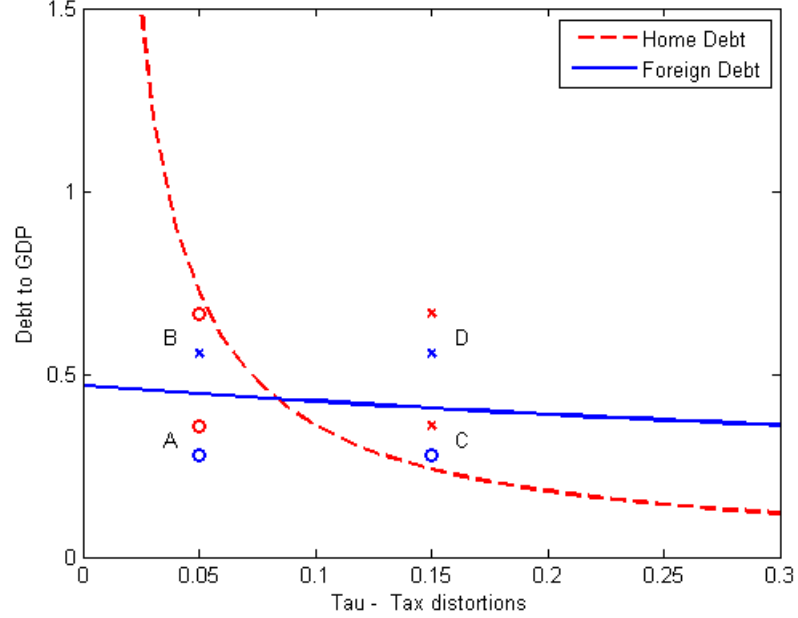
$$\frac{y_L \delta_f}{1 + \tau_L} < b_f \leq \frac{y_H \delta_f}{1 + \tau_H}, \quad (3.17)$$

$$\frac{y_H \delta_h}{\tau_H} < b_h \leq \frac{y_L \delta_h}{\tau_L}, \quad (3.18)$$

where the inequalities in (3.17) correspond to conditions 1) and 2) respectively, and the inequalities in (3.18) correspond to conditions 3) and 4) respectively. How these conditions translate into a mapping between (y, τ) outcomes and repayment-default decisions can be easily understood by looking at Figure 3.2. The red (dotted) line represents the domestic default threshold, while the blue (solid) line represents the foreign default threshold. In the second period, four situations may occur. Circles show allocations for which debt is repaid, while crosses show defaults. Colors represent respective debt types (red for home, blue for foreign). A negative shock to output is shown as an increase in the debt-to-GDP ratio.

Figure 3.2 shows four possible outcomes denoted by letters A to D. Tax distortions τ are on the horizontal axis, while the vertical axis represents domestic and foreign debt-to-GDP ratios in the second period $\frac{b_h}{y_2}$ and $\frac{b_f}{y_2}$. A negative output shock is shown as a move up, and a negative taxation shock is shown as a move to the right. A) After a good output shock and a good tax distortion shock, both debts fall below the default thresholds and therefore both are repaid. B) After a bad output shock and a good tax distortion shock, foreign debt (blue cross) is above its threshold and is therefore defaulted on. However, domestic debt (red circle) is still repaid, as it falls below its threshold. C) After a good output shock but bad a tax distortion shock, the situation is the reverse of B. D) After a bad output shock and a bad tax distortion shock, both debts are above default their thresholds and are therefore defaulted on.

Figure 3.2: Default thresholds and selective defaults



3.3.2 Debt policies in the first period

In this section we solve for first-period debt issuance decisions that are consistent with the second-period default decisions described by (3.17) and (3.18) (or equivalently by Figure 3.2). In the remainder of this paper we assume a constant relative risk aversion (CRRA) instantaneous utility function for domestic agents:

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}.$$

The aim of this section is to first find a set of parameters for which foreign default is driven by the output shock and domestic default is driven by the tax distortion shock. The solution algorithm is provided in Solution algorithm. We show that this set is non-empty (see Two-period model). Second, we examine the comparative statics of an equilibrium solution.

The government chooses debt issuances b_h and b_f to maximize the lifetime

utility of domestic agents:

$$\max_{\{b_h, b_f\}} u(c_1) + \beta \mathbb{E}[u(c_2)], \quad (3.19)$$

where

$$c_1 = y_1 + \tau_1 q_h b_h - (1 + \tau_1)(g - q_f b_f),$$

$$c_2 = \begin{cases} (3.11) & \text{with prob. } \Pi_y \Pi_\tau \\ (3.12) & \text{with prob. } (1 - \Pi_y) \Pi_\tau \\ (3.13) & \text{with prob. } \Pi_y (1 - \Pi_\tau) \\ (3.14) & \text{with prob. } (1 - \Pi_y)(1 - \Pi_\tau), \end{cases}$$

subject to price schedules derived from foreign investors' zero-profit condition and domestic households' first-order condition:

$$q_f = \frac{\Pi_y}{1 + r}, \quad (3.20)$$

$$q_h = \beta \frac{\Pi_y \Pi_\tau u'(c^r) + (1 - \Pi_y) \Pi_\tau u'(c^{fd})}{u'(c_1)}. \quad (3.21)$$

Debt issuances must obey first-order conditions given by:

$$(b_h :)(\tau_L - \tau_1)q_h = \tau_1 b_h \frac{\partial q_h}{\partial b_h}, \quad (3.22)$$

$$(b_f :)u'(c) \left((1 + \tau_1) q_f + \tau_1 b_h \frac{\partial q_h}{\partial b_h} \right) = \beta (1 + \tau_L) \left(\Pi_y \Pi_\tau u'(c^r) + \Pi_y (1 - \Pi_\tau) u'(c^{hd}) \right). \quad (3.23)$$

Comparative statics reveal that this two-period environment can account for two empirically observed facts. First, that the share of foreign investors is negatively correlated with interest rates; and second, that the share of domestic investors is positively correlated with the total public debt of the economy (see for example Andritzky (2012)). We document these findings graphically in Two-period model (graphical solutions).

Now that the trade-offs behind our model have been described in detail, we can turn to quantitative analysis of an infinite-horizon version of the model.

3.4 Quantitative analysis

We build an incomplete-markets model in which the government has limited commitment. Let time be indexed by $t = 0, 1, 2, \dots$. The economy has an exogenous stochastic stream of income $y_t \in \mathbb{Y}$, which is a Markov process. At each time t the government has to cover a fixed exogenous stream of government expenditure g_t .

In each period t the government decides either to repay or default on outstanding foreign and domestic debt. When the government chooses to default, the economy suffers from output penalties and is excluded from borrowing on the market where default happened for a random number of periods. We allow the expected exclusion durations and output costs to differ between types of default.

3.4.1 Households

Households are identical and risk averse. Their utility is given by:

$$\sum_{t=0}^{\infty} \beta^t \mathbb{E}_0 [u(c_t)],$$

where β is the discount factor, c is consumption and $u(c)$ is increasing and strictly concave. Households are allowed to save using domestically issued government bonds b_h . They take bond discount prices and taxes as given. They face an intra-temporal budget constraint, which differs depending on the government's decision to default on either of the two bonds.

If the government repays both domestic and foreign debt, households' budget constraint is the following:

$$c^r = y - T(1 + \tau) + b_h - q_h b'_h, \quad (3.24)$$

where b_h is the amount of domestic debt owed and repaid by the government to households, b'_h is the new issuance of government domestic debt (household savings), q_h is the domestic bond's discount price, T is the amount of lump-sum taxes and τ is the distortion imposed by taxation.⁴

⁴Whenever taxes are negative, the household budget constraint yields $c^r = y - T(1 - \tau) + b_h - q_h b'_h$, so that rebates are distortionary and distortion does not increase the amount of resources when taxes are negative. The same is true for the selective and total default cases.

If the government defaults on foreign debt, households are still allowed to save in the domestic market. However, foreign default induces output costs and affects the endogenous price of domestic bonds:

$$c^{fd} = y(1 - \delta_f) - T(1 + \tau) + b_h - q_h^{fd} b'_h. \quad (3.25)$$

In the case of domestic default, the government maintains foreign borrowing, but the domestic debt market is closed:

$$c^{hd} = y(1 - \delta_h) - T(1 + \tau). \quad (3.26)$$

Similarly, in the case of simultaneous domestic and foreign default, which we will refer to as total default:

$$c^{td} = y(1 - \delta_f)(1 - \delta_h) - T(1 + \tau). \quad (3.27)$$

3.4.2 Foreign investors

Foreigners are risk neutral investors with deep pockets and access to international credit markets, where they can save and borrow at a constant interest rate r . When lending resources to the government they account for the possibility of default and break even in expected terms, therefore their policy can be summarized as:

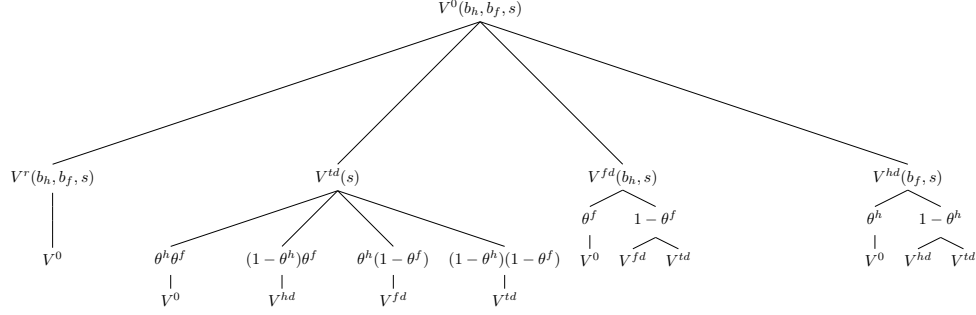
$$q_f = \frac{(1 - \Delta^f)}{1 + r},$$

where q_f is the discount price of government bonds issued with foreign investors and Δ^f is the probability of foreign default.

3.4.3 Recursive equilibrium

We define a recursive equilibrium in which domestic households, foreign investors and the government act sequentially and the government acts with discretion. The aggregate state of the economy $S = (b_h, b_f, s)$ is given by two endogenous debts b_h, b_f and two exogenous processes for income and tax distortions $s = (y, \tau)$.

Figure 3.3: Government decision tree



When both markets are open (V^0), the government can decide to repay both debts (V^r), default on both debts (V^{td}), repay only domestic debt (V^{fd}) or repay only foreign debt (V^{hd}). Subsequent possible choices are depicted on the lower levels of the decision tree.

Every period, the government decides whether to repay its two outstanding debts, default on domestic debt, default on foreign debt or default on both:

$$V^0(b_h, b_f, s) = \max\{V^r(b_h, b_f, s), V^{fd}(b_h, s), V^{hd}(b_f, s), V^{td}(s)\} \quad (3.28)$$

The government's repayment decision is summarized by two default indicators $d_f \in \{0, 1\}$ and $d_h \in \{0, 1\}$, where $d_{i=\{h,f\}} = 0$ stand for repayment, $d_f = 1$ stands for foreign default and $d_h = 1$ for domestic default. After a default, the government is excluded from borrowing on the market and faces probability θ_h, θ_f of returning to borrowing on domestic and foreign markets respectively. The government's choices are presented graphically in Figure 3.3, where tree branches correspond from left to right to: repayment of both debts, default on both debts, default on foreign debt only and default on domestic debt only. After repayment, the government goes back to node V^0 . After any type of default, the government first draws probabilities θ^h, θ^f that one or the other market will open. Subsequent possible choices are depicted on the lower levels of the tree. (Total default has been put on the second branch due to graphical reasons.)

If the government decides to repay it solves the following problem:

$$V^r(b_h, b_f, s) = \max_{b'_h, b'_f} \left\{ u(c^r) + \beta \mathbb{E} \left\{ V^0(b'_h, b'_f, s') \right\} \right\} \quad (3.29)$$

subject to households' budget constraint (3.22), the foreign bond price schedule

$$q_f(b'_f, s) = \frac{\mathbb{E}\{1 - d'_f(b'_h, b'_f, s')\}}{1 + r} \quad (3.30)$$

and households' first-order condition:

$$q_h(b_h, b_f, b'_h, b'_f, s) = \beta \frac{\mathbb{E}\left\{\left(1 - d'_h(b'_h, b'_f, s')\right) u'\left(c'(b'_h, b'_f, s')\right)\right\}}{u'(c(b_h, b_f, s))}, \quad (3.31)$$

where, unlike for foreign bonds, the price of domestic bonds depends not only on the probability of default, but also on households' welfare both today and tomorrow, and the government budget constraint

$$T + q_h b'_h + q_f b'_f = g + b_h + b_f. \quad (3.32)$$

If the government defaults on foreign debt (and keeps servicing its domestic obligations) the economy suffers an output cost, and is allowed to return to international borrowing in the future with probability θ^f . With probability $1 - \theta^f$ the country remains only on the domestic bond market and the government can still decide to also default on domestic bonds (yielding total default). The government's problem is summarized by:

$$V^{fd}(b_h, s) = \max_{b'_h} \left\{ u(c^{fd}) + \beta \mathbb{E} \left(\theta^f V^0(0, b'_h, s') + (1 - \theta^f) \max \{ V^{fd}(b'_h, s'), V^{td}(s') \} \right) \right\} \quad (3.33)$$

subject to households' budget constraint (3.23), households' first-order condition

$$q_h^{fd}(b_h, b'_h, s) = \beta \frac{\mathbb{E}\left\{\left(1 - d^{fd'}(b'_h, s')\right) u'\left(c^{fd'}(b'_h, s')\right)\right\}}{u'(c^{fd}(b_h, s))} \quad (3.34)$$

(where the number of states is reduced relative to the repayment case, as foreign debt does not affect welfare because it is defaulted on) and the government budget constraint

$$T + q_h^{fd} b'_h = g + b_h. \quad (3.35)$$

Third, if the government decides to default on domestic debt outstanding, it remains active on international markets, comes back to domestic borrowing with

probability θ^h , can still default on foreign debt and suffers a domestic output penalty:

$$V^{hd}(b_f, s) = \max_{b'_f} \left\{ u(c^{hd}) + \beta \mathbb{E} \left(\theta^h V^0(b'_f, 0, s') + (1 - \theta^h) \max \{ V^{hd}(b'_f, s'), V^{td}(s') \} \right) \right\} \quad (3.36)$$

subject to households' budget constraint (3.24), the foreign bond price schedule

$$q_f^{hd}(b'_f, s) = \frac{\mathbb{E} \{ 1 - d^{hd'}(b'_f, s') \}}{1 + r} \quad (3.37)$$

and the government budget constraint

$$T + q_f^{hd} b'_f = g + b_f. \quad (3.38)$$

Lastly, at any given time the government can decide to pursue total default. The economy suffers output penalties for both domestic and foreign default, and the government comes back to international and domestic borrowing with probabilities θ^f and θ^h respectively. The government's problem is summarized by:

$$V^{td}(s) = u(c^{td}) + \beta \mathbb{E} \left[\theta^f \theta^h V^0(0, 0, s) + \theta^f (1 - \theta^h) V^{hd}(0, s') + (1 - \theta^f) \theta^h V^{fd}(0, s') + (1 - \theta^f)(1 - \theta^h) V^{fd}(s') \right] \quad (3.39)$$

subject to households' budget constraint (3.25) and the government budget constraint

$$T = g. \quad (3.40)$$

Now that actions and optimization problems are defined for each actor in the economy, we can define the equilibrium:

Definition 1. *Recursive equilibrium in this economy is (i) the set of prices in repayment periods for domestic bonds $q_h(b_h, b_f, s)$ and foreign bonds $q_f(b_h, b_f, s)$ and the set of prices in partial default periods $q_h^{fd}(b_h, s)$ and $q_f^{hd}(b_f, s)$; (ii) government debt policies in repayment periods $b'_h(b_h, b_f, s)$ and $b'_f(b_h, b_f, s)$ and in partial default periods $b_h^{fd'}(b_h, s)$ and $b_f^{hd'}(b_f, s)$; and (iii) government default schedules in repayment periods $d_h(b_h, b_f, s)$ and $d_f(b_h, b_f, s)$ and in partial default periods $d_h^{fd}(b_h, s)$*

and $d_f^{hd}(b_f, s)$ such that:

- 1) Taking as given domestic bond price schedules d_h and d_h^{fd} and government domestic debt issuances b_h' and $b_h^{fd'}$, households' consumption c^r and c^{fd} satisfy households' budget constraints and first-order conditions.
- 2) Taking as given government foreign default schedules d_f and d_f^{hd} , prices q_f and q_f^{hd} are consistent with foreign investors' expected zero profits.
- 3) Taking as given prices q_h, q_f, q_h^{fd} and q_f^{hd} , the government's default schedules d_h, d_f, d_h^{fd} and d_f^{hd} and debt policies b_h', b_f', b_h^{fd} and b_f^{hd} solve the government's optimization problem.
- 4) Government bond and tax policies and default schedules satisfy the government budget constraint.

3.4.4 Calibration

To solve the model numerically, we need to assume specific functional forms and assign parameters. Table 3.1 represents the parameters, which are selected directly from data. We assume the CRRA utility function with a risk aversion coefficient σ equal to two. The risk-free interest rate r is set to 1.7%, which is the average yearly interest rate of a five-year US Treasury bond during this time period. These parameters are common values used in the real business cycle and default literature. We calibrate the $AR(1)$ stochastic process for output, based on the series of Argentinian GDP:

$$\log(y_t) = \rho_y \log(y_{t-1}) + u_t, \quad (3.41)$$

where $u_t \sim \mathcal{N}(0, \epsilon_y)$.

The government faces two types of costs upon default. The output cost is assumed to be asymmetric as in Arellano (2008):

$$y_t^{def} = \min\{y_t, \gamma y\}, \quad (3.42)$$

where y is the mean of the output process and γ takes one of three values for domestic, foreign and total default respectively. The cost function implies that default is more costly with a high output realization. The level of government

Table 3.1: Parameters selected directly

Parameter	Value	Source
Risk-free interest rate	$r = 1.7\%$	5-year US bond yearly yield
Risk aversion	$\sigma = 2$	Standard in literature
Persistence of output	$\rho_y = 0.945$	Argentina 1993–2001
Std. dev. of output	$\epsilon_y = 0.025$	Argentina 1993–2001
Government expenditure	$g/y = 0.25$	Argentina 1993–2001
Re-entry to foreign market	$\theta_f = 0.22$	4.6 yrs. exclusion (R&R 2011b)
Re-entry to domestic market	$\theta_h = 0.5$	2 yrs. exclusion (R&R 2011b)
Low tax distortion	$\tau_l = 0.01$	Assumed

expenditure is set to be the average Argentinian government expenditure of 25% of GDP for the period 1993–2011. This number is not substantially different from the cross-country average of 31% for developing countries. Based on Reinhart and Rogoff (2011b) dataset we calculate the median length of domestic default to be 2.5 years and that of foreign default to be 4.6 years.⁵ This estimate is slightly low in comparison with the usual average exclusion period of 7.5 years for Argentina usually applied in default literature.⁶ Our process of tax distortions is of a reduced form and cannot be directly taken to data, therefore we make two additional assumptions. First, we assume symmetry in the process (switching states from high to low and from low to high happens with the same probability). Second, we assume taxes in the good state to be almost non-distortionary. However, τ_L cannot be zero (as discussed in Proposition 5) as it would make domestic debt riskless and thereby prevent the algorithm from converging.

After choosing eight parameters directly, we are left with six parameters to be calibrated. Table 3.2 summarizes the parameters and moments that we match. We use Reinhart and Rogoff’s dataset to calculate frequencies of different types

⁵Calculated as the median of averages of defaulting countries

⁶Gelos et al. (2011) measure exclusion as the years between default and the date of the next issuance of public and publicly guaranteed bonds or syndicated loans.

Table 3.2: Parameters selected by matching moments

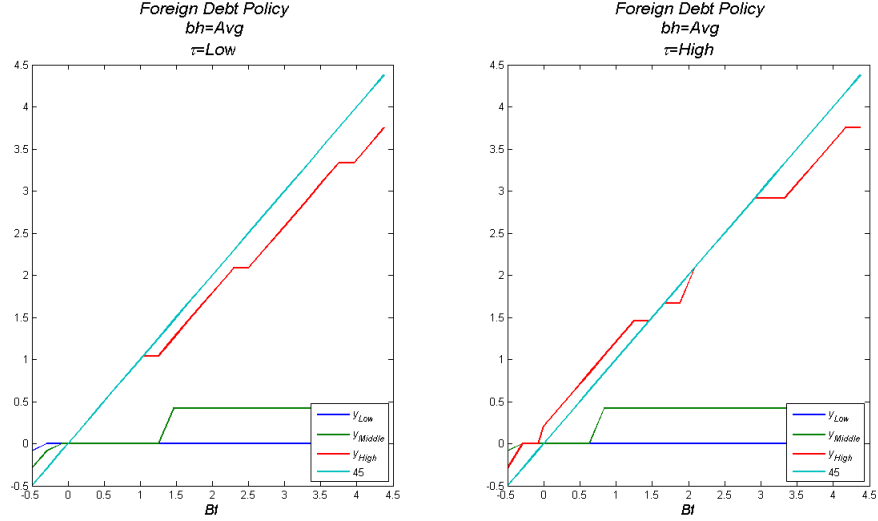
Parameter	Value	Target
Discount factor	$\beta = 0.95$	Debt service to GDP 5.53%
Foreign default output cost	$\gamma_f = 0.97$	F-default frequency 3.5%
Domestic default output cost	$\gamma_h = 0.91$	Output drop after D-default
High tax distortion	$\tau_h = 0.1$	D-debt to GDP 24.8%
High distortion persistence	$\pi_{hh} = 0.7$	D-default freq. 2.5%
Low distortion persistence	$\pi_{ll} = 0.7$	Symmetric $\pi_{hh} = \pi_{ll}$

of default, periods of market exclusion and drops in output after different types of default. As in previous literature, we calibrate the discount factor to target a debt service expenditures-to-GDP ratio of 5.53%. The foreign output cost γ^f is calibrated to match the frequency of foreign defaults in Argentina in the last 210 years. Then, we set γ^h such that the output drop after domestic default is on average three times higher than after foreign default (as documented by Reinhart and Rogoff). The persistence of distortion states are assumed to be symmetric and are set to match the frequency of domestic defaults in Argentina in the last 210 years.

Unfortunately, Reinhart and Rogoff's dataset does not report debt composition. Therefore, to calculate debt-to-GDP ratios, we employ the dataset of Panizza (2008), who constructs his data based on the legal definition, which is consistent with Reinhart and Rogoff (2011b). We try to match the domestic debt-to-GDP ratio in Argentina of 24.8%, although the model is not quite able to match this particular moment closely.

3.4.5 Simulation results

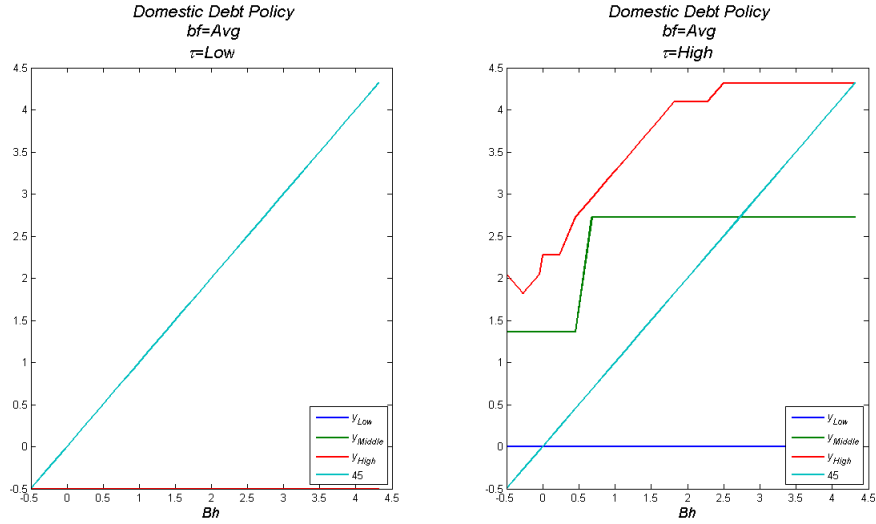
In this section we analyze default policies, debt policies and equilibrium prices in the calibrated model. Next we examine the quantitative performance of the model against the data. We describe the algorithm for solving the model numerically in Solution algorithm. Both default and debt policies are four-dimensional

Figure 3.4: Foreign debt policies given $b_h = 1.8$ 

objects, as the state space for the economy consists of two endogenous (domestic and foreign debt) and two exogenous (output and tax distortions) states. For each variable of interest, we compare policies for different levels of the same type of debt, keeping the value of the second type of debt constant.

The most interesting findings of the model are revealed by Figures 3.4 and 3.5. Figure 3.4 plots debt policies for foreign debt given that outstanding domestic debt is positive $b_h = 1.8$. Foreign debt policies are similar to those found in other quantitative models of sovereign default. The country accumulates foreign debt when output is high due to low interest rates. Interest rates are low as a result of the default set being decreasing in y . Also, the government accumulates more debt when the economy suffers from high tax distortions. This is explained by the fact that the government avoids using distortionary taxation and instead finances its expenditures via both foreign and domestic (as we shall see) debt.

Figure 3.5 plots policies for domestic debt. When tax distortions are low (left panel), the government finances its expenditures in full via taxation for any level of debt outstanding. This is the situation in which raising taxes comes at the lowest cost for the economy. In fact, the government is building up assets on the domestic debt market (optimal domestic debt is the negative corner solution) in order to

Figure 3.5: Domestic debt policies given $b_f = 1.8$ 

be able to accommodate more debt movements in the future, when distortions may be high. When tax distortions are high (right panel) and output is low, the government is in a state of default and no trade is taking place on domestic debt markets. When output is middle or high, the government employs a “gambling for redemption” policy. It finds it optimal to always increase the stock of domestic debt up to the point where it reaches endogenous debt constraints. Thus, the government is piling up domestic debt in the hope that it will be able to repay all of it with taxes, should the low-distortion day arrive. Whenever this day happens, the government repays its debt in full. If this day does not come, the government is forced to default on its domestic debt obligations.

Figures 3.6 and 3.7 plot repayment and default policies in debt–output space. White stands for repayment, light gray for foreign default, dark gray for domestic default and black for total default. We can see that the repayment–default trade-off for foreign debt is mostly driven by the output process, while tax distortions do not matter. On the other hand, the default area for domestic debt is much bigger for the high tax distortion than for the low tax distortion scenario. Also, as in both cases we set the second type of debt to zero, we cannot observe total default.

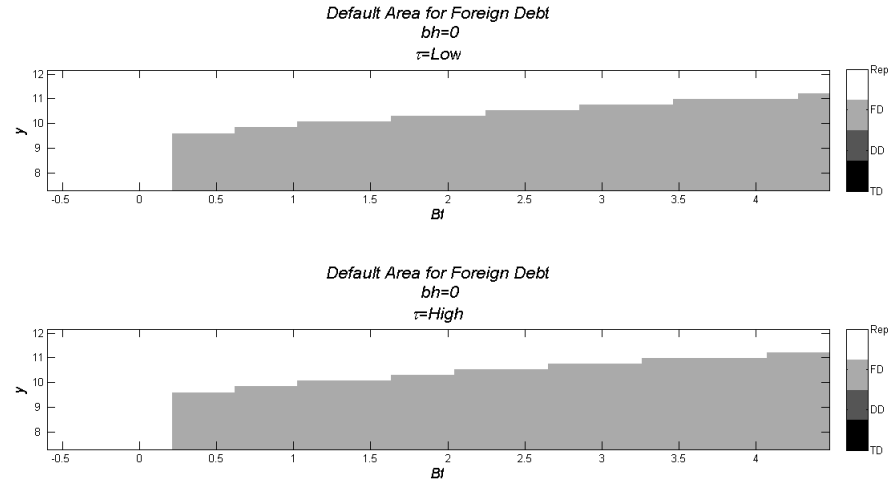
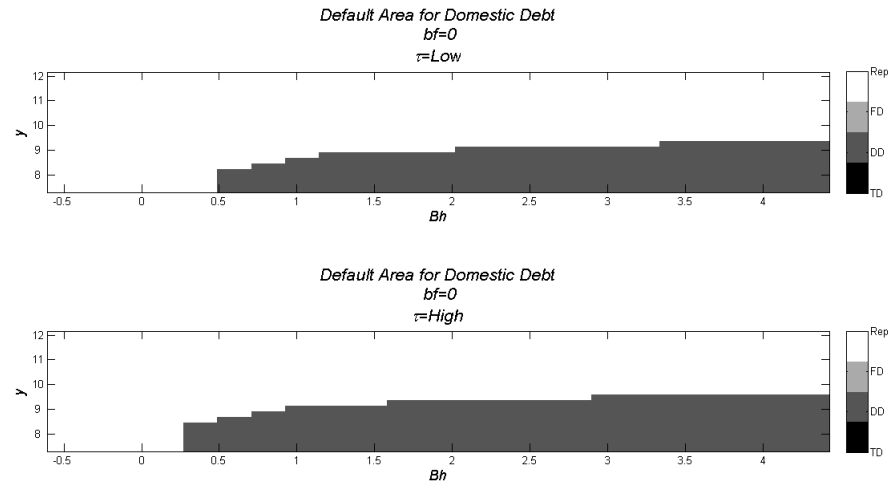
Figure 3.6: Default sets for foreign debt given $b_h = 0$ Figure 3.7: Default sets for domestic debt given $b_f = 0$ 

Table 3.3: Cyclical properties

	Data (Argentina)	Model	Arellano (2008)
Foreign default frequency	3.5%	3.5%	3%
Domestic default frequency	2.5%	5.6%	x
Total default frequency	1.5%	0.3%	x
Average foreign spread	12.67 <i>pp</i>	8.9 <i>pp</i>	3.58 <i>pp</i>
Average domestic spread	x	15.5 <i>pp</i>	x
Foreign debt-to-GDP	17.22%	3.7%	5.95%
Domestic debt-to-GDP	24.78%	13.7%	x
Consumption std./ Output std.	1.098	1.088	1.098

To assess the performance of the model, we simulate 1,000 paths from the model, each with length 10,000, and burn the first 1,000 simulations of each path. Then we compare the resulting business cycle statistics with the corresponding statistics from the data. Table 3.3 shows that the results for the benchmark calibration are in line with the data. Our model performs well in many dimensions. The model replicates reasonably high debts levels and at the same time reasonably low default probabilities. It predicts that consumption is more volatile than output, and that net exports are strongly countercyclical.⁷

It is worth stressing once again that the two shocks have opposite effects on the economy. While the tax distortion shock has a substantial impact on domestic debt accumulation, it has a mild impact on foreign debt accumulation. The opposite is true for the output shock.

⁷See Neumeyer and Perri (2005)

3.5 Secondary markets and haircuts

With the introduction of unconventional monetary policies during the Great Recession, secondary sovereign debt markets have attracted increasing interest among economists. In this section we return to the two-period model setting from Section 3.3 to study how secondary markets affect the government's incentives to repay or to default. We will introduce secondary markets in the second period. Secondary markets open after nature selects the output and taxation shocks. Therefore all participants in the market have perfect foresight of what the government will do (repay or default) if no trade in assets takes place on secondary markets.

3.5.1 Setup

The starting points for the discussion are Propositions 4, 5 and 6, where we have established that with costly tax enforcement there exist finite default thresholds for both foreign and domestic debt, and that both debts can be risky due to the stochastic nature of output and taxation distortions. There are four possible outcomes of the model in the moment at which secondary markets open, which are summarized in Figure 3.2. When either both debts are repaid (situation A) or defaulted on (situation D), the workings of the secondary markets would not change the final outcome. Therefore our discussion will focus on selective foreign default (situation B).⁸ Under situation B in the second period output is low $y_2 = y_L$ and tax distortions are low $\tau_2 = \tau_L$.

First we will summarize our assumptions about what is happening in the economy at the moment the secondary markets open, and we introduce some notation. As long as default costs are positive, there are positive amounts of both foreign and domestic debts outstanding: b_f and b_h . Each debt has its respective default threshold which we derive from (3.15) and (3.16) and denote in levels: \bar{B}_f and \bar{B}_h .

After a good shock to taxation and a bad shock to output, foreign debt is above its default threshold $b_f > \bar{B}_f$ but domestic debt lies below its default threshold

⁸Selective domestic default, situation C, is its mirror image.

$b_h < \bar{B}_h$. As foreigners know they will be defaulted upon, they are willing to sell their claims in the secondary market. As domestic investors know there is still some room for an increase in repayable domestic debt, they are willing to buy them. Bonds in the secondary market sell at discount price q^{SM} .

For the sake of consistency with the model we keep track of domestic debt outstanding. However, our analysis is also valid for the case when $b_h = 0$, as in Broner et al. (2010). Therefore we can see this section as a generalization of their work in which we allow for costly enforcement.⁹ As we shall see, what matters for creating repayment incentives through secondary markets is not the level of home or foreign debt outstanding, but the relative difference between above-the-threshold foreign holdings $b_f - \bar{B}_f$ and below-the-threshold domestic accommodation space $\bar{B}_h - b_h$. We will call the former expression “defaultable foreign debt overhang” and the latter “domestic debt accommodation space”.

We formulate this situation as a noncooperative game. There are three types of agents: domestic investors, foreign investors and the government. The outcome of the game is determined by two crucial conditions: whether every type of investors can coordinate, and whether total debt inherited from the first period within the sum of default thresholds $b_f + b_h \leq \bar{B}_h + \bar{B}_f$.

First, both populations of investors take simultaneous decisions on the amounts supplied and demanded in the secondary market given the secondary-market discount price and beliefs about the government’s subsequent action (default or repay). After trades have taken place, the secondary market closes and the government decides to either repay or default on foreign and domestic investors.

Foreign investors’ strategy space is the quantity of bonds supplied in the secondary market:

$$s^f = \{b_f^{SM}(q^{SM}, d_f^{SM})\}, \quad x_f \in [0, b_f].$$

Domestic investors’ strategy space is quantity demanded in the secondary market:

$$s^h = \{b_h^{SM}(q^{SM}, d_h^{SM})\}.$$

⁹In Broner et al. (2010), a government with discretion wants to default on foreign debt because it faces no penalties upon default $\delta^f = 0$. In our analysis, foreign default is due to the government’s discretionary behavior with $\delta^f > 0$ and an unfortunate output shock.

The government's strategy space consists of two decisions (repay or default) on the two markets:

$$s^g = \{d_f^{SM} \times d_h^{SM}\}, \quad d_f^{SM} \in \{0, 1\}, \quad d_h^{SM} \in \{0, 1\}.$$

Given strategies of the three players we can define payoffs for each player. Because of risk neutrality of the foreign investor her payoff is defined as her consumption in the second period and is the function of her decision (quantity supplied on the secondary market b_f^{SM}), domestic investor's decision (quantity demanded on the secondary market b_h^{SM}) and the government's decision (to default or repay foreign debt d_f^{SM}):

$$U^f = \begin{cases} (b_f - b_f^{SM}) + q^{SM} b_f^{SM} & \text{if } d_f^{SM} = 0 \text{ and } b_f^{SM} = b_h^{SM} \\ q^{SM} b_f^{SM} & \text{if } d_f^{SM} = 1 \text{ and } b_f^{SM} = b_h^{SM} \\ b_f & \text{if } d_f^{SM} = 0 \text{ and } b_f^{SM} \neq b_h^{SM} \\ 0 & \text{if } d_f^{SM} = 1 \text{ and } b_f^{SM} \neq b_h^{SM} \end{cases} \quad (3.43)$$

where b_f is the amount of government bonds that foreign investors hold from the first period and q^{SM} is the discount price of bonds on the secondary market in the second period. The first two cases of equation 3.41 refer to a situation when demands meets supply and there is trade in bonds on the secondary market. The last two cases refer to a situation when there is no trade on the secondary market. On the other hand first and third case of equation 3.41 describe payoffs to a foreign investor when the government repays foreign debt, whereas second and fourth case describe payoffs to a foreign investor when the governments defaults on foreign debt.

Similarly, we define the payoff for the domestic investor. Her payoff differs from foreign investor's payoff mainly due to risk aversion. The payoff of the domestic investor is the utility from consumption ($u(c)$) as defined in equation 3.19) in the second period after secondary market closes. The domestic investor decides on the quantity demanded in the secondary market b_h^{SM} taking the supply of bonds from foreign investors b_f^{SM} and the government decision to default or repay domestic

debt d_h^{SM} as given:

$$U^h = \begin{cases} u(y_2 + b_h + b_h^{SM}(1 - q^{SM}) - T_2(1 + \tau_2)) & \text{if } d_h^{SM} = 0 \text{ and } b_f^{SM} = b_h^{SM} \\ u(y_2 - q^{SM}b_h^{SM} - T_2(1 + \tau_2)) & \text{if } d_h^{SM} = 1 \text{ and } b_f^{SM} = b_h^{SM} \\ u(y_2 + b_h - T_2(1 + \tau_2)) & \text{if } d_h^{SM} = 0 \text{ and } b_f^{SM} \neq b_h^{SM} \\ u(y_2 - T_2(1 + \tau_2)) & \text{if } d_h^{SM} = 1 \text{ and } b_f^{SM} \neq b_h^{SM} \end{cases} \quad (3.44)$$

Finally, the government moves after the secondary market closes. The government decides whether to default or repay both debts d_h^{SM}, d_f^{SM} taking b_h^{SM} and b_f^{SM} as given and its payoff is defined by (3.42). The government decision boils down to two default thresholds policies as shown in derivations (3.11)-(3.16). These policies, given the trade on the secondary market in the second period, translate to:

$$d_h^{SM}(b_h^{SM}, b_f^{SM}) = \begin{cases} 0 & \text{if } b_h + b_h^{SM} \leq \bar{B}_h \\ 1 & \text{if } b_h + b_h^{SM} > \bar{B}_h \end{cases} \quad (3.45)$$

$$d_f^{SM}(b_h^{SM}, b_f^{SM}) = \begin{cases} 0 & \text{if } b_f - b_f^{SM} \leq \bar{B}_f \\ 1 & \text{if } b_f - b_f^{SM} > \bar{B}_f \end{cases} \quad (3.46)$$

A Nash equilibrium of this game is the triplet of strategies $\{s_{f*}, s_{h*}, s_{g*}\}$ for which quantity demanded b_h^{SM} equals quantity supplied b_f^{SM} given market clearing price q^{SM*} and beliefs of investors are consistent with the government decisions d_h^{SM*}, d_f^{SM*} :

$$b^{SM*} = b_f^{SM}(q^{SM*}, d_f^{SM*}) = b_h^{SM}(q^{SM*}, d_h^{SM*}) \quad (3.47)$$

$$d_h^{SM*} = d_h^{SM}(b_h^{SM*}, b_f^{SM*}) \quad (3.48)$$

$$d_f^{SM*} = d_f^{SM}(b_h^{SM*}, b_f^{SM*}) \quad (3.49)$$

We split our analysis into two parts. In the first, we analyze the situation when foreign debt overhang is greater than domestic debt accommodation space $(b_f - \bar{B}_f) > (\bar{B}_h - b_h)$. That is, in order to be repaid, foreign investors have to sell more bonds than domestic investors can accommodate and still be repaid. It is

thus impossible that both groups be repaid after the secondary market closes. In the second part, we analyze the reverse situation, when foreign debt overhang is smaller than domestic debt accommodation space ($b_f - \bar{B}_f) < (\bar{B}_h - b_h)$. In this situation domestic investors can safely buy what foreign investors need to supply in order to be repaid. In theory, secondary markets could allow both groups of investors to be repaid.

We will look for Nash equilibria in pure strategies with continuous strategy sets. The precise outcomes of the model will depend on the assumptions we make about the possibility of investor coordination and of voluntary debt haircuts. In terms of investor coordination, we consider two different cases. First, we consider the case in which the set of investors is a continuum (infinite number of investors, each investor has size zero). Second, we modify this assumption and introduce a finite number of investors (each investor has size ϵ). This theoretical notion has a very intuitive interpretation in our game. By assumption, a zero-size investor does not internalize the effects of her individual decision on aggregate action of the set of investors of her class (domestic or foreign), whereas an ϵ -size investors does. If there are externalities in this game (and we shall see that indeed externalities arise) then an ϵ -size investor internalizes them. Therefore it is equivalent to say that zero-size investors cannot coordinate their actions while ϵ -size investors can coordinate. For each of the two parts (foreign debt overhang dominates, domestic accommodation space dominates) we will analyze four cases, when each set of investors either can or cannot coordinate.

The second important assumption is either forbidding or allowing free disposal. When free disposal is forbidden, the amount of bonds issued must be equal to the amount of bonds claimed. When free disposal is allowed, each investor can voluntarily burn some of her bonds, so the amount of bonds claimed can be lower than the amount of bonds issued. Free disposal also has a very intuitive interpretation in our example. When free disposal is allowed and exercised, we can think of this as a voluntary debt haircut.

Table 3.4 gives a brief summary of the results of secondary markets and haircuts, when foreign debt overhang dominates. We are initially, before the secondary

Table 3.4: Secondary markets and haircuts

Foreign debt overhang is greater

Investors		Foreign	
Domestic	Coordinate	Coordinate	Don't coordinate
		FD (δ^f); $b^{SM} \in (\bar{B}_h - b_h, b_f - \bar{B}_f)$	FD (δ^f); $b^{SM} = \bar{B}_h - b_h$
		SM reduce welfare	SM reduce welfare
		Haircut restores repayment	No Haircut equilibrium
	Don't coordinate	DD (δ^h); $b^{SM} = b_f - \bar{B}_f$	TD ($\delta^f + \delta^h$);
	coordinate	SM likely increase welfare	SM reduce welfare
		Haircut restores repayment	No Haircut equilibrium

market opens, in situation B: the economy would suffer the foreign default cost δ^f and the amount of domestic debt b^h needs to be either rolled over or repaid by distortionary taxation. If trade on the secondary market does not alter the outcome in terms of default of the primary market (*the first row* of Table 3.4), any trade on secondary market is undesirable from the welfare point of view, because it either increases the risk of default or induces dead-weight losses of distortionary taxation. If trade on the secondary market alters the outcome in terms of default of the primary market (*the second row* of Table 3.4), the welfare analysis is ambiguous in some cases. However, we can provide intuition for some cases. First, if both domestic and foreign investors are infinitesimal, the economy suffers the output loss $(1 - \delta^f)(1 - \delta^h)y$ upon total default instead of the output loss $(1 - \delta^f)y$ upon foreign default, but both debts are set to zero. However, if it was desirable to have total default from the welfare point of view, total default would happen on primary market. Second, if foreign investors are ϵ -size and domestic investors are infinitesimal, the economy suffers the output loss $(1 - \delta^h)y$ upon domestic default instead of the output loss $(1 - \delta^f)y$ upon foreign default and a substantial reduction of foreign debt.

Table 3.5 gives a brief summary of the results of secondary markets and haircuts, when domestic accommodation space dominates. We are initially, before the secondary market opens, in situation B: the economy would suffer the foreign default cost δ^f and the amount of domestic debt b^h needs to be either rolled over or

Table 3.5: Secondary markets and haircuts

Domestic accommodation space is greater

Investors		Foreign	
		Coordinate	Don't coordinate
Domestic	Coordinate	SM restore repayment on both markets	
	Don't coordinate	No Haircut equilibrium	

repaid by distortionary taxation. Trade on the secondary market restores repayment of both debt. if both domestic and foreign investors are infinitesimal, the economy suffers the output loss $(1 - \delta^f)(1 - \delta^h)y$ upon total default instead of the output loss $(1 - \delta^f)y$ upon foreign default, but both debts are set to zero. However, if it was desirable to have total default from the welfare point of view, total default would happen on primary market. Second, if foreign investors are ϵ -size and domestic investors are infinitesimal, the economy suffers the output loss $(1 - \delta^h)y$ upon domestic default instead of the output loss $(1 - \delta^f)y$ upon foreign default and a substantial reduction of foreign debt.

3.5.2 Equilibria when foreign debt overhang dominates

Proposition 7. *If both domestic and foreign investors are infinitesimal (cannot coordinate) and the defaultable foreign debt overhang is greater than the domestic debt accommodation space:*

- Nash equilibrium is indeterminate and degenerate.*
- $b^{SM} \in (\bar{B}_h - b_h, b_f - \bar{B}_f), \quad q^{SM} = 0.$
- Both debts are defaulted on: $d_h^{SM} = 1, \quad d_f^{SM} = 1.$*

Proof. See Appendix 3.7.4. □

The only Nash equilibrium under this specification of the game is indeterminate and occurs at a discount price equal to zero. This specification of the game suffers

from a well known equilibrium existence problem due to discontinuous payoffs (see Dasgupta and Maskin (1986a) and Dasgupta and Maskin (1986b)). Because of discontinuous payoffs, the best response functions of both investors do not cross at any positive price, which is demonstrated in the proof of the proposition.

Proposition 4 shows then that under certain circumstances secondary markets do not help create incentives for repayment of government debt, when those incentives are absent on the primary market. Moreover, as the equilibrium is indeterminate and degenerate, the outcome of secondary market trade is uncertain. This result may shed some light on why, in turbulent times, secondary markets may cease to function. Russia in 1998 effectively defaulted on its obligations towards households but repaid its obligations to firms. Why there was no significant re-trade of bonds between households and firms on the secondary market remains an open question, but this proposition may provide some intuition.

We investigate this result further in altering the assumptions that neither domestic nor foreign investors can coordinate their actions. We formalize this idea by relaxing the assumption of each investor being zero-measure. Instead we assume that the measure of each investor is $\epsilon > 0$, so that the economy is populated by $\frac{1}{\epsilon}$ investors.

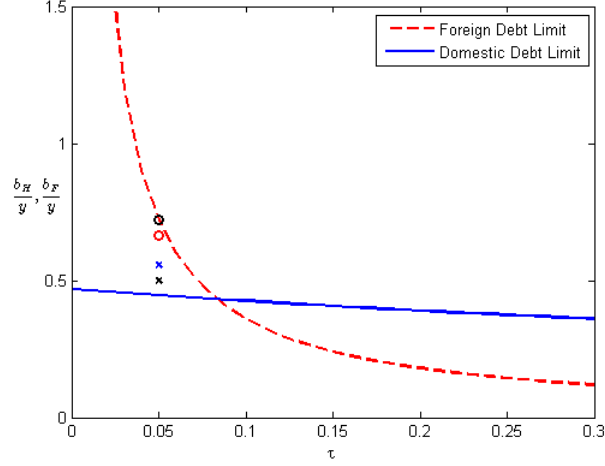
Proposition 8. *If domestic investors are ϵ -size (are able to coordinate) and the defaultable foreign debt overhang is greater than the domestic debt accommodation space:*

- a. *Nash equilibrium in pure strategies is indeterminate but yields a unique allocation.*
- b. $b^{SM} = \bar{B}_h - b_h, \quad q^{SM} = 0.$
- c. *Domestic debt is repaid $d_h^{SM} = 0$ and foreign debt is defaulted on $d_f^{SM} = 1.$*

Proof. See Appendix 3.7.4. □

The result in Proposition 5 is graphically depicted in Figure 3.8. The red circle and the blue cross show the situation before trade on the secondary market. With trade, domestic investors increase their holdings up to their default threshold and are repaid (black circle). Foreign investors decrease their holdings, but are

Figure 3.8: Trade in secondary markets improves welfare (Proposition 5)



nevertheless defaulted on (black cross) as their bond holdings are still above the default threshold. Note that the equilibrium in Proposition 5 holds when foreign investors both can and cannot coordinate. The ability to coordinate among domestic investors is not only a sufficient condition to sustain repayment incentives on the domestic market, but also allow domestic investors to capture the whole surplus generated by trade on the secondary market ($q^{SM} = 0$). Hence, trade on the secondary market does not affect welfare of foreign investors.

Lastly, let us study the reverse situation. Now foreign investors can coordinate and domestic investors are all zero-measure.

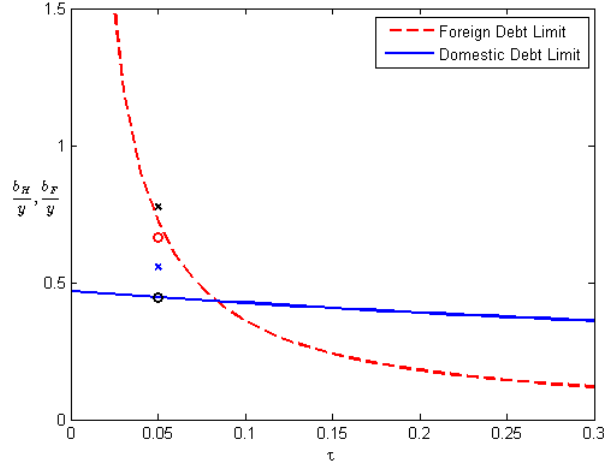
Proposition 9. *If foreign investors are ϵ -size (are able to coordinate), domestic investors are infinitesimal (cannot coordinate) and the defaultable foreign debt overhang is greater than the domestic debt accommodation space:*

- a. *Nash equilibrium in pure strategies is unique and degenerate.*
- b. $b^{SM} = b_f - \bar{B}_f$, $q^{SM} = 0$.
- c. *Domestic debt is defaulted on $d_h^{SM} = 1$ and foreign debt is repaid $d_f^{SM} = 0$.*

Proof. See Appendix 3.7.4. □

Proposition 9 shows an interesting result. Under the mix of unfavorable circum-

Figure 3.9: Trade in secondary markets destroys welfare (Proposition 6)



stances for domestic investors (a low accommodation space relative to the foreign debt overhang, and a lack of domestic coordination while foreign investors can coordinate), introducing a secondary market reverses the selective default result that would otherwise occur on the primary market.

This situation is shown in Figure 3.9. Again, the blue cross and the red circle stand for the situation before secondary markets open (foreign default and domestic repayment). Now foreign investors are able to re-trade their defaultable debt overhang to home investors, and are repaid by the government (black circle). Domestic investors exceed the domestic debt default threshold and are defaulted on by the government (black cross). Instead of defaulting on its foreign obligations, the government defaults on domestic debt holdings, and foreign obligations are repaid.

Proposition 10. *If foreign investors are allowed free disposal and are ϵ -size (can coordinate):*

- a. *The game is reduced to two players: foreign investors and the government.*
- b. *Equilibrium is unique.*
- c. *$b_f - \bar{B}_f$ is freely disposed of.*
- d. *Both debts are repaid: $d_h = 0$ and $d_f = 0$.*

Proof. See Appendix 3.7.4. □

Proposition 11. *If foreign investors are allowed free disposal and are infinitesimal (cannot coordinate), the Nash equilibrium in pure strategies does not exist.*

Proof. See Appendix 3.7.4. □

Propositions 10 and 11 show that voluntary haircuts would occur only when foreign investors are able to coordinate. This is because without coordination each investor has incentives to deviate from the haircut allocation and freely dispose less than $b_f - \bar{B}_f$, not expecting this would change the government's decision. After a voluntary haircut, foreign debt is repaid. Domestic debt is unaffected and also repaid. Interestingly, voluntary haircuts increase welfare and restore repayment incentives in situations when secondary markets may fail to deliver a well-behaved equilibrium (Proposition 9).

Results in this section may shed some light on the Greek government debt crisis, when in 2012 private investors agreed to a voluntary haircut while the trade of government bonds on secondary markets was negligible.

3.5.3 Equilibria when domestic accommodation space dominates

In this part we analyze the situation in which the foreign debt overhang is smaller than the domestic debt accommodation space $(b_f + b_h) < (\bar{B}_h + \bar{B}_f)$.

Proposition 12. *If domestic investors can accommodate all of the defaultable foreign debt overhang:*

- a. *Nash equilibrium is indeterminate (but well-behaved).*
- b. $b^{SM} \in (b_f - \bar{B}_f, \bar{B}_h - b_h)$, $q^{SM} = 1$.
- c. *Both debts are repaid: $d_h^{SM} = 0$ and $d_f^{SM} = 0$.*

Proof. See Appendix 3.7.4. □

The result in Proposition 12 is similar to Broner et al. (2010). In their paper, before the secondary market opens the government wants to default on foreign

investors (and domestic debt is zero). In the secondary market, foreigners re-trade all of their holdings to domestic investors, and the government repays in full to domestic investors. Here, foreign investors only re-trade the amount above their default threshold (in Broner et al. (2010) this threshold is zero), but it is enough to restore repayment on the foreign market. Domestic investors increase their holdings, but are still below the default threshold (in the cited paper this threshold is infinity) and are therefore also repaid by the government. The necessary condition for secondary markets to restore repayment on both markets when tax enforcement is costly is that the foreign debt overhang is smaller than the domestic debt accommodation space $(b_f - \bar{B}_f) < (\bar{B}_h - b_h)$.

This result affects the workings of the primary market, as it turns risky foreign debt into riskless debt. Therefore the discount price on the primary market is $q_f = \frac{1}{1+r}$.

The aim of this section is to show that the effects of secondary markets for government bonds are ambiguous in the situation where either domestic or foreign debt would otherwise be defaulted on. This section by no means exhausts the topic. What this section proves is that strengthening the role and efficiency of secondary markets is not a remedy that can automatically solve the sovereign risk problem. We find that the equilibria are dependent on underlying conditions, such as investors' coordination abilities and the relative size of demand and supply of bonds. Clearly more research, both empirical and theoretical, is warranted on the workings of secondary markets during sovereign risk crisis.

3.6 Conclusions

We develop a model of sovereign debt issuance on international and domestic markets, and of selective defaults. By adding domestic investors we introduce a new level of heterogeneity to a standard model of strategic sovereign default. Our model is capable of replicating selective default frequencies and business cycle statistics, and we show that including two types of investors brings the model closer to the data, as it was suggested by Aguiar and Amador (2014). Our model is a

useful tool to study how the fractions of investors in public debt arise endogenously in an equilibrium, and how the composition of debt is correlated with spreads and the total debt. Our model shows that although foreign debt is more valuable and can in principle be used to smooth both output and taxation shocks, the government would still use domestic debt to smooth the domestic taxation shock. In a world with two uncorrelated shocks (output and taxation), two types of debt (foreign and domestic) are issued, and selective defaults arise endogenously (as we observe in the data).

On the positive side, we provide a theory of the role of secondary sovereign debt markets in restoring repayment incentives. Trade in secondary markets can restore the government's repayment incentives when the supply of defaultable bonds from foreigners is low compared to demand from domestic investors. However, when the supply of defaultable bonds is high (compared to demand), then secondary markets cannot sustain repayment on both markets. If domestic investors are able to coordinate, then trade in secondary markets can be welfare-improving for both sides. Otherwise, if domestic investors cannot coordinate, then it is uncertain whether any trade would occur on secondary markets.

On the other hand, if foreign investors are able to coordinate then they will be willing to accept a voluntary haircut on the eve of foreign default. This would restore debt repayment on the foreign market. In the absence of coordination, foreign investors will never accept haircuts and foreign debt will be always defaulted on. In particular situations when secondary markets fail to improve the allocation, a voluntary haircut does, and *vice versa*. Our results shed some light on the Greek government debt crisis, when in 2012 private investors agreed to a voluntary haircut while trade in government bonds on secondary markets was negligible.

How investors' coordination may arise endogenously is an interesting and important issue for further research. However, as investors' coordination improves the allocation and welfare outcomes, we hypothesize that within a group each investor has incentives to defect on coordination and free-ride on the coordinating majority. Instances of this behavior have been seen in recent default episodes, especially prior to the 2014 Argentinian default.

Bibliography

- AGUIAR, M. AND M. AMADOR (2014): “Sovereign Debt,” in *Handbook of International Economics Vol 4*, North-Holland, 647–687.
- ANDRITZKY, J. R. (2012): “Government Bonds and their Investors,” IMF Working Papers 12/158, International Monetary Fund.
- ARELLANO, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *American Economic Review*, 98, 690–712.
- BRONER, F., A. MARTIN, AND J. VENTURA (2010): “Sovereign Risk and Secondary Markets,” *American Economic Review*, 100, 1523–55.
- COOPER, R., H. KEMPF, AND D. PELED (2008): “Is It Is Or Is It Ain’t My Obligation? Regional Debt In A Fiscal Federation,” *International Economic Review*, 49, 1469–1504.
- DASGUPTA, P. AND E. MASKIN (1986a): “The Existence of Equilibrium in Discontinuous Economic Games, I: Theory,” *Review of Economic Studies*, 53, 1–26.
- (1986b): “The Existence of Equilibrium in Discontinuous Economic Games, II: Applications,” *Review of Economic Studies*, 53, 27–41.
- DELL’ERBA, S., R. HAUSMANN, AND U. PANIZZA (2013): “Debt Levels, Debt Composition, and Sovereign Spreads in Emerging and Advanced Economies,” *Oxford Review of Economic Policy*, 29, 518–547.

- D'ERASMO, P. AND E. G. MENDOZA (2013): "Distributional Incentives in an Equilibrium Model of Domestic Sovereign Default," NBER Working Papers 19477, National Bureau of Economic Research, Inc.
- EATON, J. AND M. GERSOVITZ (1981): "Debt with Potential Repudiation: Theoretical and Empirical Analysis," *Review of Economic Studies*, 48, 289–309.
- ERCE, A. (2012): "Selective Sovereign Defaults," Tech. rep.
- GELOS, R. G., R. SAHAY, AND G. SANDLERIS (2011): "Sovereign borrowing by developing countries: What determines market access?" *Journal of International Economics*, 83, 243–254.
- MERLER, S. AND J. PISANI-FERRY (2012): "Who's afraid of sovereign bonds?" Policy Contributions 695, Bruegel.
- NEUMEYER, P. A. AND F. PERRI (2005): "Business cycles in emerging economies: the role of interest rates," *Journal of Monetary Economics*, 52, 345–380.
- PANIZZA, U. (2008): "Domestic And External Public Debt In Developing Countries," Tech. rep.
- POUZO, D. AND I. PRESNO (2014): "Optimal Taxation with Endogenous Default under Incomplete Markets," *Working paper*.
- REINHART, C. M. AND K. S. ROGOFF (2011a): "The Forgotten History of Domestic Debt," *Economic Journal*, 121, 319–350.
- (2011b): "From Financial Crash to Debt Crisis," *American Economic Review*, 101, 1676–1706.
- VASISHTHA, G. (2010): "Domestic versus External Borrowing and Fiscal Policy in Emerging Markets," *Review of International Economics*, 18, 1058–1074.

3.7 Appendix

3.7.1 Two-period model

A. Algorithm

We solve for the government's optimal domestic and foreign debt policies in the first period following these steps:

1. Assuming that (3.17) and (3.18) are satisfied in the second period, we write the government's problem as (3.19).
2. The solution to the problem is then a set of two first-order conditions (3.3.2) and (3.3.2) and pricing rules (3.20) and (3.21).
3. We pick a set of parameters and solve (3.19) numerically.
4. We confirm that the resulting policy functions b_f , b_h and equilibrium prices q_f , q_h satisfy conditions (3.17)–(3.18), and therefore that expectations in (3.19) are consistent in equilibrium.
5. We vary one parameter at a time within a range where (3.17)–(3.18) are satisfied to derive comparative statics.

B. Parametrization

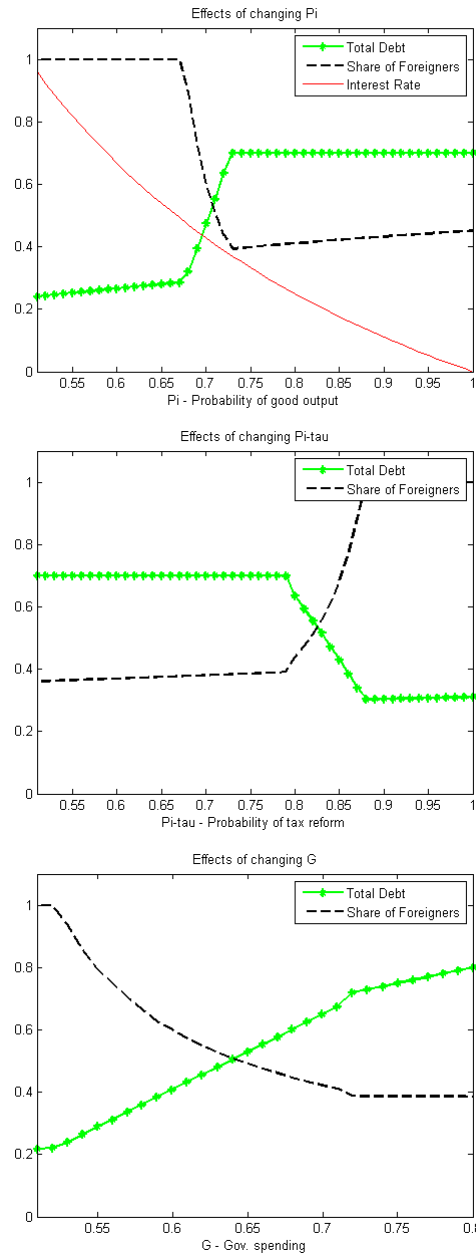
Parameter	Value	Range	Description
y	1		Output today / High output tomorrow
σ	1		Risk aversion of Home agents
Π_y	0.72	[0.5 , 1]	Probability of high output
Π_τ	0.8	[0.5 , 1]	Probability of low tax distortion
g	0.7	[0.5 , 0.8]	Gov. expenditure
y_L	0.5	[0.1 0.7]	Low output
τ_1	0.1	[0 , 0.2]	Tax distortions today
τ_H	0.15	[0.1 , 0.2]	Tax distortions tomorrow (high)
τ_L	0.05	[0 , 0.15]	Tax distortions tomorrow (low)
δ_f	0.65	[0.42 , 0.87]	Output cost of Foreign default
δ_h	0.05	[0 , 0.15]	Output cost of Home default
r	0.00		Risk-free interest rate

3.7.2 Solution algorithm

1. Guess price schedules p_f^0 and p_h^0 .
2. Calculate consumption in autarky c^{aut} and value of permanent autarky V^{aut} .
3. Guess four value functions V^{00} , V^{0fd} , V^{0hd} and V^{0td} using V^{aut} .
4. Calculate optimal policies b_f and b_h in repayment given V^{00} as continuation value and prices.
5. Calculate value of repayment V^r given optimal policies and continuation value.
6. Repeat steps 4 and 5 for foreign default and domestic default to obtain V^{1fd} and V^{1hd} .
7. Calculate value of total default V^{1td} given V^{1fd} and V^{1hd} and V^{00} .
8. Derive optimal default policies d comparing four value functions V^r , V^{1fd} , V^{1hd} , V^{1td} at each grid point $\{b_f, b_h, y, \tau\}$.

-
9. Derive new value function V^{10} as maximum of four value functions used in previous step at each grid point.
 10. Substitute $V^{00} = V^{10}$.
 11. Repeat steps 3–9 until convergence in value function.
 12. Given optimal default policies d calculate prices of foreign and domestic debt q_f^1 and q_h^1 at each grid point using pricing rules (3.29) and (3.28).
 13. Update prices $q_f^0 = \alpha^f q_f^0 + (1 - \alpha^f) q_f^1$ and $q_h^0 = \alpha^h q_h^0 + (1 - \alpha^h) q_h^1$.
 14. Repeat steps 1–13 until convergence in prices.

3.7.3 Two-period model (graphical solutions)



3.7.4 Secondary market: proposition proofs

Proof of Proposition 7

We study situation B depicted in the Figure 3.2: outcome on the primary market yields foreign default (FD) and domestic repayment (DR). Before engaging in the trade on the secondary markets, both types of investors (domestic and foreign) need to form expectations about the government's decision to default. There are three possible outcomes: foreign default and domestic repayment (FD, DR); foreign repayment and domestic default (FR, DD) and total default (FD, DD). The repayment of both debts is ruled out by the fact that total debt is greater than the total default limit $b_f + b_h > \bar{B}_h + \bar{B}_f$.

There are two thresholds related to the volume traded in the secondary markets b^{SM} . If the traded volume lies below lower threshold $b^{SM} \leq \bar{B}_h - b_h$, domestic debt is repaid, while foreign debt is defaulted. If the traded volume is between lower and upper threshold $\bar{B}_h - b_h < b^{SM} < b_f - \bar{B}_f - b_h$, both debts are defaulted; if the traded volume lies above upper threshold $b^{SM} \geq \bar{B}_h - b_h$ domestic debt is defaulted, while foreign debt is repaid.

We draw three best-response correspondences: for foreign investors (*the solid red line*), for domestic investors (*the dashed blue line*) and for the government in a single graph with the amount of trade b^{SM} on the horizontal axis and the price q^{SM} of debt on the vertical axis. *The brown shaded area* represents the area of trade, in which the expectations are consistent with the government decision: in the first panel the shaded area represents the amounts traded for which the government will choose (FD, DR), in the second panel the shaded area represents the amounts traded for which the government will choose (FD, DD) and in the third panel shaded area represents the amount traded for which the government will choose (FR, DD). A crossing of the two best best response correspondences, which lies within shaded area represents a Nash equilibrium in pure strategies (secondary markets clear and expectations of investors are consistent with the government's decision). As can be seen in Figure 3.10, the only outcome that is consistent is the total default. In this case the price on the secondary markets is zero, but the

amount of trade is undetermined. \square

Proof of Proposition 8

We follow a similar procedure to 3.7.4. However, there is one substantial difference. If domestic investors are ϵ -size, they internalize the effects of their actions on the government's decision. Since $b_f + b_h > \bar{B}_h + \bar{B}_f$ at least one type of debt will be defaulted. The outcome on the primary market yields domestic repayment. Since each domestic investor is ϵ -size, she will never demand any amount that exceeds domestic default threshold, as this will unambiguously decrease her payoff. Hence $x_h \leq \bar{B}_h - b_h$. In this, domestic investors can effectively insure domestic repayment.

Therefore, the only possible outcome that is consistent in equilibrium is foreign default and domestic repayment (FD,DR). Contrary to 3.7.4 we can narrow our considerations and study only one game when both types of investors expect (FD, DR). The best response function of single foreign investor depends on whether she is: (a) zero-size, (b) ϵ -size and expecting domestic investors to be zero-size (uninformed foreign investor) or (c) ϵ -size and knows that domestic investors are ϵ -size (informed foreign investor). In Figure 3.11 we draw best response functions for the three cases.

If foreign investors are infinitesimal (panel (a)) they do not coordinate and each of them wants to sell all of her debt holdings (b_f) as long as price on the secondary markets is positive. Hence, there exist a unique equilibrium where at zero price q^{SM} the maximum possible amount of debt $b^{SM} = \bar{B}_h - b_h$, that insures domestic repayment, is traded. Secondly, when foreign investors are ϵ -size but uninformed (panel (b)), they coordinate their supply on the amount that exceeds $b_f - \bar{B}_h$ (that will insure they are repaid by the government in the primary market) as long as price is positive. Similarly, there exist a unique equilibrium where at zero price q^{SM} the maximum possible amount of debt $b^{SM} = \bar{B}_h - b_h$, that insures domestic repayment, is traded. Thirdly, if foreign investors are ϵ -size and informed that domestic investors coordinate (panel (c)), the traded amount remains $b^{SM} = \bar{B}_h - b_h$, but the price is undetermined $q^{SM} \in [0, 1]$. \square

Proof of Proposition 9

We consider the opposite case to 3.7.4. Foreign investors are ϵ -size, they internalize the effect of their actions on the government decision, while domestic investors are infinitesimal. Since $b_f + b_h > \bar{B}_h + \bar{B}_f$, at least the one type of debt will be defaulted by the government. Hence, foreign investors coordinate to insure foreign repayment (FR). The only possible outcome that is consistent in equilibrium is foreign repayment and domestic default (FR,DD). In Figure 3.12 we draw the best response functions for this case. There exists a unique equilibrium, where at zero price q^{SM} the minimum possible amount $b^{SM} = b_f - \bar{B}_f$, that insures foreign repayment, is traded. \square

Proof of Propositions 10 and 11

We consider an alternative way (compared to Proposition 9) to bring foreign debt b_f down (weakly) below its default threshold \bar{B}_f . Foreign investors are ϵ -size, they internalize the effect of their actions on the government decision. As shown in the previous proposition, engaging in the secondary market does not bring direct benefit for foreign investors, but it might restore foreign repayment under very specific circumstances. The option of free disposal plays a similar role. It does not bring a direct benefit, but it insures foreign repayment as long as foreign investors can coordinate on the minimum amount of a haircut $b_f - \bar{B}_f$. The free disposal requires only the coordination between foreign investors. \square

Proof of Proposition 12

We study situation B depicted in the Figure 3.2: outcome on the primary market yields foreign default (FD) and domestic repayment (DR). Before engaging in the trade on the secondary markets, both types of investors (domestic and foreign) need to form expectations about the government's decision to default. There are three possible outcomes: foreign default and domestic repayment (FD, DR); foreign repayment and domestic default (FR, DD) and total default (FD, DD). The repayment of both debts is ruled out by the fact that total debt is

greater than the total default limit $b_f + b_h > \bar{B}_h + \bar{B}_f$.

There are two thresholds related to the volume traded in the secondary markets b^{SM} . If the traded volume lies below lower threshold $b^{SM} \leq \bar{B}_h - b_h$, domestic debt is repaid, while foreign debt is defaulted. If the traded volume is between lower and upper threshold $\bar{B}_h - b_h < b^{SM} < b_f - \bar{B}_f - b_h$, both debts are defaulted; if the traded volume lies above upper threshold $b^{SM} \geq \bar{B}_h - b_h$ domestic debt is defaulted, while foreign debt is repaid.

We draw three best-response correspondences: for foreign investors (*the solid red line*), for domestic investors (*the dashed blue line*) and for the government in a single graph with the amount of trade b^{SM} on the horizontal axis and the price q^{SM} of debt on the vertical axis. *The brown shaded area* represents the area of trade, in which the expectations are consistent with the government decision. A crossing of the two best best response correspondences, which lies within shaded area represents a Nash equilibrium in pure strategies (secondary markets clear and expectations of investors are consistent with the government's decision).

In Figure 3.13 we draw best response correspondences for the case of infinitesimal foreign and infinitesimal domestic investors. Only in panel (b), depicting the case when both types of investors expect domestic repayment and foreign repayment (FR, DR), best correspondences cross in the shaded area, which means that expectations are consistent in equilibrium. In a unique equilibrium both debt are repaid, the volume volume is however undetermined $b^{SM} \in (b_f - \bar{B}_f, \bar{B}_h - b_h)$ and the price is equal to one $q^{SM} = 1$.

Since $b_f + b_h < \bar{B}_h + \bar{B}_f$, both debts can be potentially repaid, even without coordination. Coordination of foreign/domestic investors reinforces foreign/domestic repayment. Interestingly, coordination does not change the outcome. In Figure 3.14 we plot best response correspondences of domestic and foreign investors together with government's optimal default decision in single graphs for three remaining cases: (a) ϵ -size foreign and infinitesimal domestic investors, (b) infinitesimal foreign and ϵ -size domestic investors and (c) ϵ -size foreign and ϵ -size domestic investors. For each case best response correspondences differ slightly, but an equilibrium is the same across all three cases.

Finally, let us consider two other cases, in which one class of investors consists of ϵ -size agents, whereas the other of infinitesimal agents. As shown in Propositions 8 and 9 the class that consists of agents of ϵ -size has an advantage, as they can coordinate on their most favorable outcome. We want to check, whether in the case studied here, the side with an advantage can coordinate on their most favorable outcome.

In panel (a) of Figure 3.15 we plot best response correspondences when foreign investors are ϵ -size (have an advantage) and domestic investors are infinitesimal. Both investors expect government to default on domestic debt and repay the foreign debt (FR, DD). Even though foreign investors might potentially coordinate on any level of debt, for example $\bar{B}_h - b_h$, they would not do so. This volume and the price $q^{SM} = 0$ cannot be an equilibrium, because there is a profitable deviation for each foreign investors to reduce her traded volume down to $b_f - \bar{B}_f$ and therefore to secure domestic repayment. Foreign repayment and domestic default (FR, DD) cannot be sustained as an equilibrium.

In panel (b) of Figure 3.15 we plot best response correspondences when domestic investors are ϵ -size (have an advantage) and foreign investors are infinitesimal. Both investors expect government to default on foreign debt and repay the domestic debt (FD, DR). Even though domestic investors might potentially coordinate on any level of debt, for example $b_f - \bar{B}_f$, they would not do so. This volume and the price $q^{SM} = 0$ cannot be an equilibrium, because there is a profitable deviation for each domestic investors to always increase the traded volume, up until the point, where it meets supply from foreign investors if $\bar{B}_h - b_h$ and therefore to secure foreign repayment. Foreign default and domestic repayment (FD, DR) cannot be sustained as an equilibrium.

□

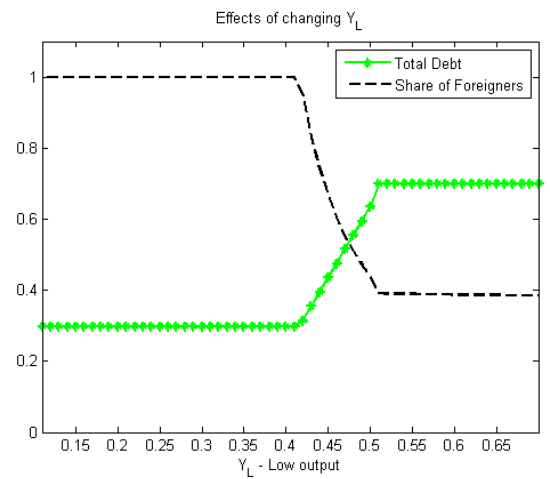
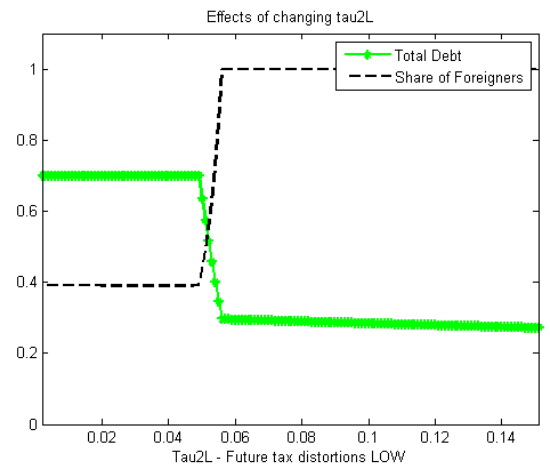
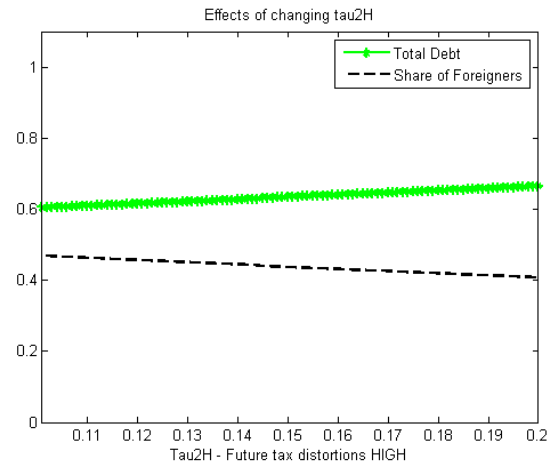


Figure 3.10: Best response functions for infinitesimal investors

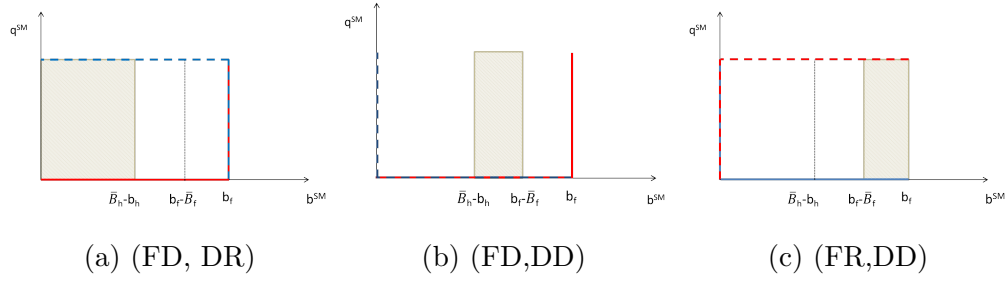
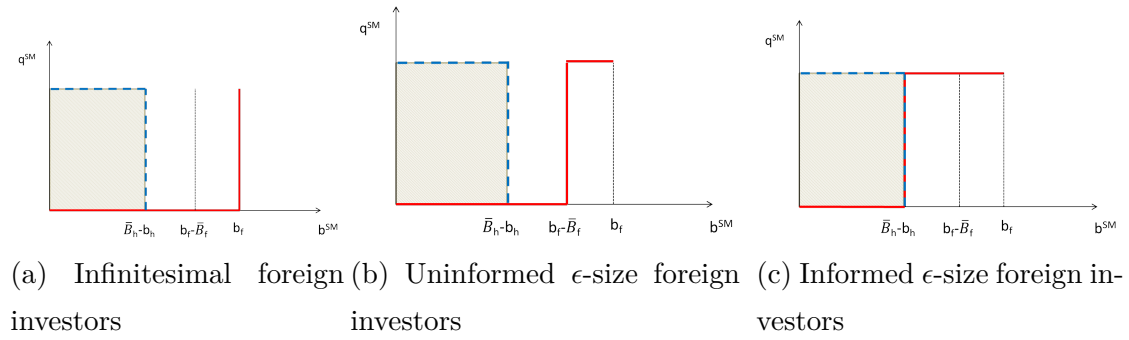
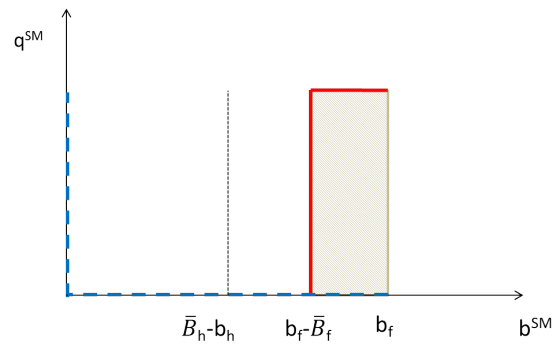
Figure 3.11: Best response functions for ϵ -size domestic investorsFigure 3.12: Best response functions for ϵ -size foreign and infinitesimal domestic investors

Figure 3.13: Best response functions for infinitesimal foreign and infinitesimal domestic investors

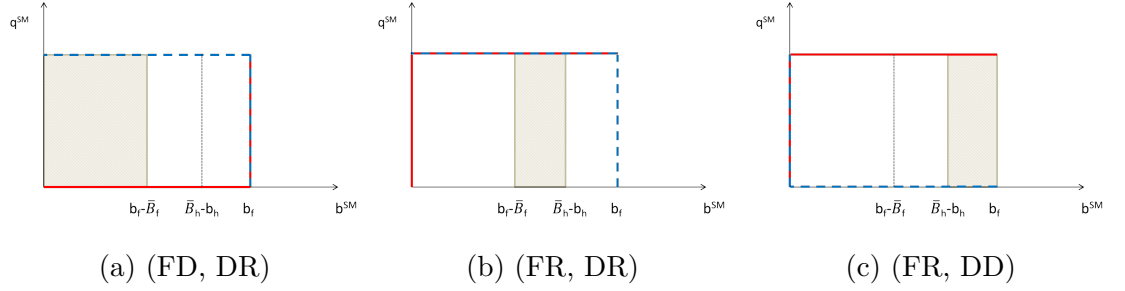


Figure 3.14: Best response functions for ϵ -size investors

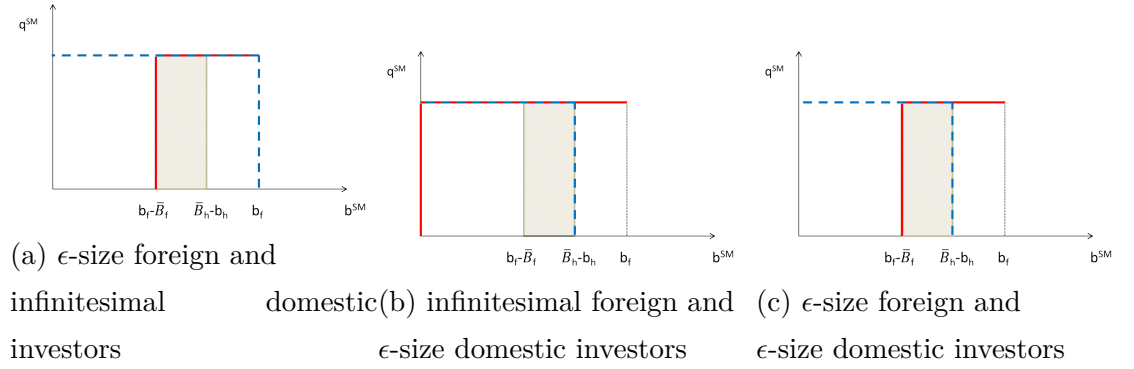
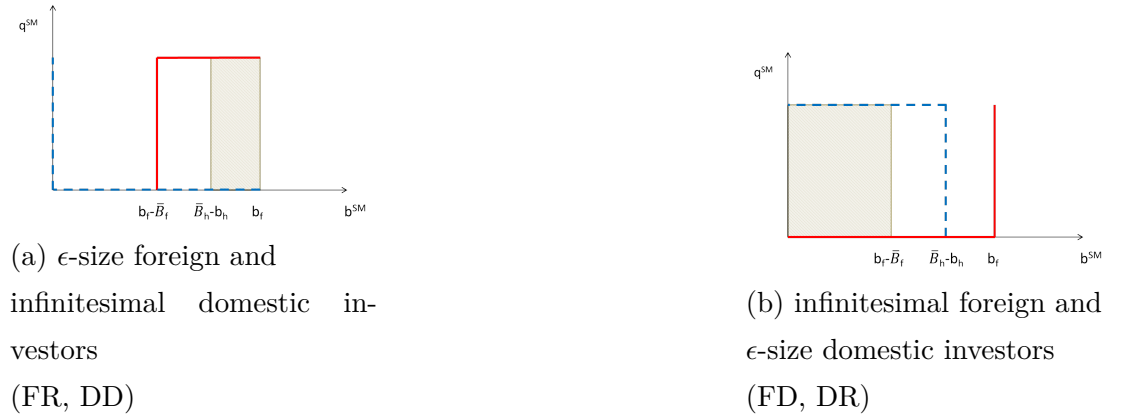


Figure 3.15: Best response functions with one-sided advantage



Bibliography

- ACEMOGLU, D. (2002): “Directed Technical Change,” *Review of Economic Studies*, 69, 781–809.
- ACEMOGLU, D. AND D. AUTOR (2011): *Skills, Tasks and Technologies: Implications for Employment and Earnings*, Elsevier, vol. 4 of *Handbook of Labor Economics*, chap. 12, 1043–1171.
- AGHION, P. (2002): “Schumpeterian Growth Theory and the Dynamics of Income Inequality,” *Econometrica*, 70, 855–882.
- AGUIAR, M. AND M. AMADOR (2014): “Sovereign Debt,” in *Handbook of International Economics Vol 4*, North-Holland, 647–687.
- ANDRITZKY, J. R. (2012): “Government Bonds and their Investors,” IMF Working Papers 12/158, International Monetary Fund.
- ARELLANO, C. (2008): “Default Risk and Income Fluctuations in Emerging Economies,” *American Economic Review*, 98, 690–712.
- AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): “Trends in U.S. Wage Inequality: Revising the Revisionists,” *The Review of Economics and Statistics*, 90, 300–323.
- BANERJEE, A. V. AND A. F. NEWMAN (1991): “Risk-Bearing and the Theory of Income Distribution,” *Review of Economic Studies*, 58, 211–35.
- BAUER, C. AND J. V. R. MORA (2014): “The Joint Determination of TFP and Financial Sector Size,” Working paper.

- BAUMOL, W. J. (1967): *American Economic Review*, 57, 415–426.
- (2001): *Paradox of the services: exploding costs, persistent demand*, vol. 1 of *The growth of service industries : the paradox of exploding costs and persistent demand*, Edward Elgar, Cheltenham.
- BECKER, G. S. (1964): *Human capital: A theoretical and empirical analysis*, 11271, National Bureau of Economic Research.
- (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81, 813–46.
- BERKES, E., U. PANIZZA, AND J.-L. ARCAND (2012): “Too Much Finance?” IMF Working Papers 12/161, International Monetary Fund.
- BIENZ, C. AND J. HIRSCH (2011): “The Dynamics of Venture Capital Contracts,” *Review of Finance*, 16, 157–195.
- BLANCHARD, O. J. (1985): “Debt, Deficits, and Finite Horizons,” *Journal of Political Economy*, 93, 223–47.
- BLANCHARD, O. J. AND S. FISCHER (1989): *Lectures on Macroeconomics*, vol. 1 of *MIT Press Books*, The MIT Press.
- BOLTON, P., T. SANTOS, AND J. A. SCHEINKMAN (2011): “Cream skinning in financial markets,” NBER Working Papers 16804, National Bureau of Economic Research, Inc.
- BRONER, F., A. MARTIN, AND J. VENTURA (2010): “Sovereign Risk and Secondary Markets,” *American Economic Review*, 100, 1523–55.
- BRUNNERMEIER, M. K., T. M. EISENBACH, AND Y. SANNIKOV (2012): “Macroeconomics with Financial Frictions: A Survey,” NBER Working Papers 18102, National Bureau of Economic Research, Inc.
- BUERA, F. J. AND J. P. KABOSKI (2009a): “Can Traditional Theories of Structural Change Fit The Data?” *Journal of the European Economic Association*, 7, 469–477.

- (2009b): “The Rise of the Service Economy,” .
- (2011): “Scale and the origins of structural change,” *Journal of Economic Theory*, 1, 1–29.
- BUERA, F. J. AND Y. SHIN (2013): “Financial Frictions and the Persistence of History: A Quantitative Exploration,” *Journal of Political Economy*, 121, 221 – 272.
- CAHUC, P. AND E. CHALLE (2012): “Produce Or Speculate? Asset Bubbles, Occupational Choice, And Efficiency,” *International Economic Review*, 53, 1105–1131.
- CAMERON, S. V. AND J. J. HECKMAN (2001): “The Dynamics of Educational Attainment for Black, Hispanic, and White Males,” *Journal of Political Economy*, 109, 455–499.
- CARD, D. AND T. LEMIEUX (2001): “Can Falling Supply Explain The Rising Return To College For Younger Men? A Cohort-Based Analysis,” *The Quarterly Journal of Economics*, 116, 705–746.
- CARDAK, B. A. (2004): “Education Choice, Endogenous Growth and Income Distribution,” *Economica*, 71, 57–81.
- CHEN, Q., I. GOLDSTEIN, AND W. JIANG (2006): “Price Informativeness and Investment Sensitivity to Stock Price,” *Review of Financial Studies*, 20, 619–650.
- CLARK, C. (1940): *The Conditions of Economic Progress*, London/New York.
- COOLEY, T. F., R. MARIMON, AND V. QUADRINI (2013): “Risky Investments with Limited Commitment,” NBER Working Papers 19594, National Bureau of Economic Research, Inc.
- COOPER, R., H. KEMPF, AND D. PELED (2008): “Is It Is Or Is It Ain’t My Obligation? Regional Debt In A Fiscal Federation,” *International Economic Review*, 49, 1469–1504.

- DASGUPTA, P. AND E. MASKIN (1986a): “The Existence of Equilibrium in Discontinuous Economic Games, I: Theory,” *Review of Economic Studies*, 53, 1–26.
- (1986b): “The Existence of Equilibrium in Discontinuous Economic Games, II: Applications,” *Review of Economic Studies*, 53, 27–41.
- DE FRAJA, G. (2002): “The Design of Optimal Education Policies,” *Review of Economic Studies*, 69, 437–66.
- DELL’ERBA, S., R. HAUSMANN, AND U. PANIZZA (2013): “Debt Levels, Debt Composition, and Sovereign Spreads in Emerging and Advanced Economies,” *Oxford Review of Economic Policy*, 29, 518–547.
- D’ERASMO, P. AND E. G. MENDOZA (2013): “Distributional Incentives in an Equilibrium Model of Domestic Sovereign Default,” NBER Working Papers 19477, National Bureau of Economic Research, Inc.
- DURNEV, A., R. MORCK, AND B. YEUNG (2004): “Value-Enhancing Capital Budgeting and Firm-specific Stock Return Variation,” *Journal of Finance*, 59, 65–105.
- EATON, J. AND M. GERSOVITZ (1981): “Debt with Potential Repudiation: Theoretical and Empirical Analysis,” *Review of Economic Studies*, 48, 289–309.
- ECHEVARRIA, C. (1997): “Changes in Sectoral Composition Associated with Economic Growth,” *International Economic Review*, 38, 431–52.
- EICHER, T. (2001): “Inequality and growth: the dual role of human capital in development,” *Journal of Development Economics*, 66, 173–197.
- ERCE, A. (2012): “Selective Sovereign Defaults,” Tech. rep.
- FREDRIKSEN, K. B. (2012): “Less Income Inequality and More Growth - Are they Compatible? Part 6. The Distribution of Wealth,” OECD Economics Department Working Papers No. 929, OECD.

- FREIXAS, X. AND J.-C. ROCHET (2008): *Microeconomics of Banking, 2nd Edition*, MIT Press Books, The MIT Press.
- GALOR, O. AND O. MOAV (2004): “From Physical to Human Capital Accumulation: Inequality and the Process of Development,” *Review of Economic Studies*, 71, 1001–1026.
- GELOS, R. G., R. SAHAY, AND G. SANDLERIS (2011): “Sovereign borrowing by developing countries: What determines market access?” *Journal of International Economics*, 83, 243–254.
- GENNAIOLI, N., A. SHLEIFER, AND R. W. VISHNY (2013): “Finance and the Preservation of Wealth,” NBER Working Papers 19117, National Bureau of Economic Research, Inc.
- GERTLER, M. AND N. KİYOTAKI (2010): “Financial Intermediation and Credit Policy in Business Cycle Analysis,” in *Handbook of Monetary Economics*, ed. by B. M. Friedman and M. Woodford, Elsevier, vol. 3 of *Handbook of Monetary Economics*, chap. 11, 547–599.
- GLOMM, G. AND B. RAVIKUMAR (1992): “Public versus Private Investment in Human Capital Endogenous Growth and Income Inequality,” *Journal of Political Economy*, 100, 813–34.
- GOLDIN, C. AND L. F. KATZ (2007): “The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005,” NBER Working Papers 12984, National Bureau of Economic Research, Inc.
- (2008): “Transitions: Career and Family Life Cycles of the Educational Elite,” *American Economic Review*, 98, 363–69.
- GOLDSMITH, R. (1995): *Financial Structure and Development*, vol. 1, Yale University Press.
- GORDON, P. (2000): “The Americas,” *Global Finance*, 14.

- GOURINCHAS, P.-O. AND J. A. PARKER (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89.
- GREENAWAY, D. AND M. HAYNES (2003): "Funding Higher Education in The UK: The Role of Fees and Loans," *Economic Journal*, 113, F150–F166.
- GREENSPAN, A. (2002): "World Finance and Risk Management," <http://www.federalreserve.gov/BoardDocs/Speeches/2002/200209253/default.htm>.
- GREENWOOD, R. AND D. SCHARFSTEIN (2013): "The Growth of Finance," *Journal of Economic Perspectives*, 27, 3–28.
- HOSIOS, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *Review of Economic Studies*, 57, 279–98.
- HSIEH, C.-T. AND P. J. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 124, 1403–1448.
- JONES, C. I. (1995): "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 103, 759–84.
- JOVANOVIC, B. (2014): "Misallocation and Growth," *American Economic Review*, 104, 1149–71.
- JOVANOVIC, B. AND B. SZENTES (2013): "On the Market for Venture Capital," *Journal of Political Economy*, 121, 493 – 527.
- KANE, T. J. (1994): "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, 102, 878–911.
- KAPLAN, S. N. AND P. STROMBERG (2003): "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts," *Review of Economic Studies*, 70, 281–315.
- KNEER, C. (2012): "The Absorption of Talent into Finance: Evidence from U.S. Banking Deregulation," *Job Market Paper of Tilburg University*, 1–28.

- KOHLBERG, E. (1976): “A model of economic growth with altruism between generations,” *Journal of Economic Theory*, 13, 1–13.
- LEVINE, R. (2005): “Finance and Growth: Theory and Evidence,” in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Elsevier, vol. 1 of *Handbook of Economic Growth*, chap. 12, 865–934.
- LEVINSOHN, J. AND A. PETRIN (2003): “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 70, 317–341.
- LEWIN, T. (2008): “College May Become Unaffordable for Most in U.S.” *The New York Times*.
- LOURY, G. C. (1981): “Intergenerational Transfers and the Distribution of Earnings,” *Econometrica*, 49, 843–67.
- LUCAS, ROBERT E, J. AND L. A. RAPPING (1969): “Real Wages, Employment, and Inflation,” *Journal of Political Economy*, 77, 721–54.
- LUCAS, R. J. (1988): “On the mechanics of economic development,” *Journal of Monetary Economics*, 22, 3–42.
- LUO, Y. (2005): “Do Insiders Learn from Outsiders? Evidence from Mergers and Acquisitions,” *The Journal of Finance*, 60, 1951–1982.
- MADDALA, G. S. AND S. WU (1999): “A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test,” *Oxford Bulletin of Economics and Statistics*, 61, 631–52.
- MADDISON, A. (2005): “Measuring And Interpreting World Economic Performance 1500-2001,” *Review of Income and Wealth*, 51, 1–35.
- MAYER-HAUG, K., S. READ, J. BRINCKMANN, N. DEW, AND D. GRICHNIK (2013): “Entrepreneurial talent and venture performance: A meta-analytic investigation of SMEs,” *Research Policy*, 1–23.

- MERLER, S. AND J. PISANI-FERRY (2012): “Who’s afraid of sovereign bonds?” Policy Contributions 695, Bruegel.
- MERTON, R. C. (1995): *The Global Financial System: A Functioning Perspective*, Harvard Business School Press.
- MGI (2011): “Big data: The next frontier for innovation, competition, and productivity,” The mckinsey global institute report, The McKinsey Global Institute.
- MICHELACCI, C. AND J. SUAREZ (2004): “Business Creation and the Stock Market,” *Review of Economic Studies*, 71, 459–481.
- MIDRIGAN, V. AND D. Y. XU (2014): “Finance and Misallocation: Evidence from Plant-Level Data,” *American Economic Review*, 104, 422–58.
- MILGROM, P. AND J. ROBERTS (1990): “Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities,” *Econometrica*, 58, 1255–77.
- MURPHY, K. M., A. SHLEIFER, AND R. W. VISHNY (1991): “The Allocation of Talent : Implications for Growth,” *The Quarterly Journal of Economics*, 106, 503–530.
- NASH, J. (1950): “The Bargaining Problem,” *Econometrica*, 18, 155–162.
- (1953): “Two-Person Cooperative Games,” *Econometrica*, 21, 128–140.
- NEUMEYER, P. A. AND F. PERRI (2005): “Business cycles in emerging economies: the role of interest rates,” *Journal of Monetary Economics*, 52, 345–380.
- NGAI, L. R. AND C. A. PISSARIDES (2007): “Structural Change in a Multisector Model of Growth,” *American Economic Review*, 97, 429–443.
- NORDHAUS, W. D. (1982): “Economic policy in the face of declining productivity growth,” *European Economic Review*, 18, 131–157.
- OBAMA, B. (2009): “The President Explains His Larger Vision on the Economy,” <http://www.whitehouse.gov/video/The-President-Explains-His-Larger-Vision-on-the-Economy>.

- OECD (2011): “OECD Science, Technology and Industry Scoreboard,” Oecd report, OECD.
- PANIZZA, U. (2008): “Domestic And External Public Debt In Developing Countries,” Tech. rep.
- PESARAN, M. H. (2007): “A simple panel unit root test in the presence of cross-section dependence,” *Journal of Applied Econometrics*, 22, 265–312.
- PHILIPPON, T. (2010): “Financiers versus Engineers: Should the Financial Sector be Taxed or Subsidized?” *American Economic Journal: Macroeconomics*, 2, 158–182.
- PHILIPPON, T. AND A. RESHEF (2012): “Wages and Human Capital in the U.S. Finance Industry: 1909–2006,” *The Quarterly Journal of Economics*, 127, 1551–1609.
- PIKETTY, T. AND G. ZUCMAN (2014): “Capital is Back: Wealth-Income Ratios in Rich Countries, 1700-2010,” *The Quarterly Journal of Economics*.
- POUZO, D. AND I. PRESNO (2014): “Optimal Taxation with Endogenous Default under Incomplete Markets,” *Working paper*.
- REINHART, C. M. AND K. S. ROGOFF (2011a): “The Forgotten History of Domestic Debt,” *Economic Journal*, 121, 319–350.
- (2011b): “From Financial Crash to Debt Crisis,” *American Economic Review*, 101, 1676–1706.
- RESTUCCIA, D. AND R. ROGERSON (2008): “Policy Distortions and Aggregate Productivity with Heterogeneous Plants,” *Review of Economic Dynamics*, 11, 707–720.
- RODRIGUEZ, S. B., J. DIAZ-GIMENEZ, V. QUADRINI, AND J.-V. RIOS-RULL (2002): “Updated facts on the U.S. distributions of earnings, income, and wealth,” *Quarterly Review*, 2–35.

- ROGERSON, R. (2008): "Structural Transformation and the Deterioration of European Labor Market Outcomes," *Journal of Political Economy*, 116, 235–259.
- ROMER, P. M. (1986): "Increasing Returns and Long-Run Growth," *The Journal of Political economy*, 94, 1002–1037.
- SAHLMAN, W. A. (1990): "The structure and governance of venture-capital organizations," *Journal of Financial Economics*, 27, 473–521.
- SANTARELLI, E., L. KLOMP, AND A. THURIK (2006): "Gibrat's Law: An Overview of the Empirical Literature," in *Entrepreneurship, Growth, and Innovation*, ed. by E. Santarelli, Springer US, vol. 12 of *International Studies in Entrepreneurship*, 41–73.
- SCHALTEGGER, C. A. AND C. GORGAS (2011): "The Evolution of Top Incomes in Switzerland over the 20th Century," CREMA Working Paper Series 2011-06, Center for Research in Economics, Management and the Arts (CREMA).
- SCHUMPETER, J. A. (1934): *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and the Business Cycle*, Transaction Publishers.
- SHU, P. (2012): "Essays on Innovation, Productivity, and Talent Allocation," *MIT Job Market Paper*.
- TOBIN, J. (1984): "On the efficiency of the financial system," *Lloyds Bank Review*, 153, 1–15.
- TOPKIS, D. M. (1998): *Supermodularity and Complementarity*, Princeton University Press, first edition ed.
- VASISHTHA, G. (2010): "Domestic versus External Borrowing and Fiscal Policy in Emerging Markets," *Review of International Economics*, 18, 1058–1074.
- VEALL, M. (2010): "Top Income Shares in Canada: Updates and Extension," *WORKING PAPER*.

- VIVES, X. (2005): “Complementarities and Games: New Developments,” *Journal of Economic Literature*, 43, 437–479.
- VOLCKER, P. (2009): “Paul Volcker: Think More Boldly,” <http://online.wsj.com/articles/SB10001424052748704825504574586330960597134>.
- WADHWA, V., B. A. RISSING, AND G. GEREFFI (2006): “Industry Trends in Engineering Offshoring,” report, Duke University.
- WALKER, I. AND Y. ZHU (2008): “The College Wage Premium and the Expansion of Higher Education in the UK,” *Scandinavian Journal of Economics*, 110, 695–709.
- WEISS, M. (2008): “Skill-biased technological change: Is there hope for the unskilled?” *Economics Letters*, 100, 439–441.
- YAARI, M. E. (1965): “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer,” *The Review of Economic Studies*, 32, 137–150.
- YANG, F. (2009): “Consumption over the Life Cycle: How Different is Housing?” *Review of Economic Dynamics*, 12, 423–443.